

THESE

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

Pour obtenir le grade de:

DOCTEUR D'UNIVERSITE

Spécialité Productique: Automatique et Informatique Industrielle

par

DELMOTTE François

Ingénieur EUDIL

ANALYSE MULTI-MODELE

Soutenue le 29 septembre 1997 devant le jury d'examen:

Président	Professeur M. Staroswiecki
Rapporteurs	Professeur A. Richard Professeur D. Meizel Professeur M. Ksouri
Examineurs	Professeur P. Borne Professeur G. Dauphin-Tanguy Docteur A. M. Desodt-Jolly Professeur D. Dubois Professeur P. Penel

Travaux réalisés sous la direction du Professeur P. Borne au Laboratoire d'Automatique et d'Informatique de Lille, à l'Ecole Centrale de Lille.

Remerciements

La recherche présentée au travers de ce document a été effectuée au Laboratoire d'Automatique et d'Informatique Industrielle de l'Ecole Centrale de Lille, sous la direction de Monsieur Pierre Borne.

Monsieur Borne m'a appris ce qu'était le métier de chercheur, et ses conseils, ainsi que la grande liberté qu'il m'a accordé, ont fait que cette recherche n'aurait pas pu se dérouler dans de meilleurs conditions. En outre je n'oublierai jamais qu'il n'a pas hésité à m'accueillir dans son équipe alors que cela n'était pas possible dans mon ancien laboratoire. Je voudrais lui présenter mes plus sincères remerciements pour tout ce qu'il a fait.

Je voudrais remercier Monsieur Marcel Staroswiecki pour avoir accepté d'être le président du jury. J'éprouve toujours un sentiment de nostalgie en repensant à ses cours, qui alliaient une grande clarté à une grande rigueur, et je suis flatté qu'il ait pris cette responsabilité.

Monsieur Mekhi Ksouri a délaissé ses importantes charges en Tunisie pour assister à la soutenance de ma thèse en tant que rapporteur. En cela, il internationalise mon jury, et je suis très honoré par sa venue.

Un membre éminent du célèbre CRAN a accepté d'être rapporteur de ma thèse. En outre il m'a dit qu'il appréciait l'humour qui parsemait ce document, ce qui est très important à mes yeux. Je voudrais remercier Monsieur Alain Richard qui m'honore de sa présence.

Monsieur Dominique Meizel m'a dit avant notre entrevue qu'il se souvenait de notre première rencontre au congrès WAC 96 il y a plus d'un an. Je suis flatté qu'un professeur de son envergure se soit remémoré cette rencontre avec un simple thésard. Je voudrais le remercier vivement pour avoir accepté d'être rapporteur de cette thèse.

Je m'enorgueillis de la présence de Monsieur Didier Dubois, qui est membre du jury. Ses travaux ont accompagné bien de mes lectures depuis mon lointain DEA sur les fonctions de croyance.

Je suis très heureux de la présence de Madame Anne-Marie Jolly-Desodt, du CAL. Celle qui a accompagné mes premiers balbutiements en recherche m'a prodigué nombre de conseils, et son rire perçant m'a toujours rendu de bonne humeur.

J'avais décidé de quitter les classes préparatoires à l'issue de ma première année de *Mathématiques spéciales*. J'aurais voulu intégrer une licence de maths, afin poursuivre une formation complète dans ce beau domaine. Cependant mon père voulait que je sois ingénieur, c'est pourquoi j'ai intégré l'EUDIL. Bien que la suite de mon parcours se soit révélée plus que réussie, j'éprouve toujours une pointe de regret de n'avoir pas poursuivi mes études en maths. C'est pourquoi je suis très touché par la présence de Monsieur Penel.

Enfin, je voudrais remercier Madame Geneviève Dauphin-Tanguy, l'égérie du LAIL. Celle qui est un peu la "maman" de tous les thésards du LAIL s'est toujours efforcée de les conseiller (moi y compris) au mieux, afin qu'ils puissent trouver un travail une fois devenu docteur.

Après les membres du jury de ma thèse, je voudrais évoquer tous ceux que j'ai côtoyé durant ces trois années.

D'abord je voudrais avoir une pensée pour tous mes camarades du LAIL et d'autres laboratoires, qui se rongent les sangs en se demandant de quoi sera fait leur avenir. Nous avons bien ri ensemble, et cela a été très positif sur la recherche. Il y avait Arnaud, Aude, Hervé (P.), Janette, Stéphane, Moncef (ceux là ont un emploi stable), et Mathieu, Pascal, Wajdi, Hervé (C), Khaled (M.), Philippe (T.), Fatime, Laurent (D.), Philippe (K.) Khaled (G.), Régis, William, Sonia, Roger, Jean-Marc, Laurent (L.), David, Patrick, Gilles, Christophe, Francine, Mélanie, Zoubert, (pardon aux mécaniciens pour mon article dans le journal il y a deux ans), Florence (L.).

Ensuite je voudrais remercier les membres de l'Ecole Centrale de Lille qui m'ont le plus aidé dans mes diverses activités: Monsieur Serniclay pour les relations de l'administration avec l'ACEC, l'association des jeunes chercheurs de l'école, Monsieur Verhelst, pour le logo de l'ACEC, Madame Bekaert, pour les recharges de photocopie, dont j'ai fait un usage immodéré, Monsieur Gallais, qui a toujours réparé les ordinateurs rebelles, Madame Tricot, pour tous mes petits problèmes administratifs, Mlle Bukowski et Mme Wignacoiurt, qui m'ont aidé à préparer mon voyage aux Etats Unis, et enfin Max, sans qui ce document n'existerait pas.

Enfin je voudrais remercier tous ceux du labo et de l'école que je n'ai pas directement mentionnés, mais qui par leur bonne humeur et leurs conseils ont joué pour moi un rôle certain durant ces années.

Table des Matières

TABLE DES MATIERES	5
INTRODUCTION GENERALE	11
DES ENSEMBLES FLOUS A LA LOGIQUE FLOUE	15
1.1 Introduction: Pourquoi les ensembles flous	16
1.2 Définition d'un ensemble flou	16
1.2.1 Origine des ensembles flous	16
1.2.2 Lien avec la théorie classique: les coupures α	17
1.3 Négation, Union, Intersection	18
1.3.1 Définitions Classiques	18
1.3.1.1 Complément	19
1.3.1.2 Intersection	19
1.3.1.3 Union	20
1.3.2 Généralisation des opérateurs avec les t-normes et t-conormes	20
1.3.2.1 Les t-normes et t-conormes	20
1.3.2.2 Exemples de t-normes et t-conormes	21
1.3.2.3 Origine des t-normes et t-conormes	25
1.3.2.4 Compatibilité avec la théorie classique	26
1.3.3 Opérateurs étranges	27
1.3.3.1 Opérateurs de maximisation et de minimisation	27
1.3.3.2 Opérateurs mi-intersection mi-union	28
1.4 Des ensembles flous à la logique floue	30
1.4.1 Projection, Produit Cartésien et Extension	31
1.4.2 Implication floue	31
1.4.3 La logique floue: Le Modus Ponens généralisé	33
1.4.3.1 Le principe	33
1.4.3.2 Compatibilité entre les opérateurs	34
1.4.3.3 Etude de l'évolution de la conclusion	35
1.4.3.4 Exemple de calcul d'une implication	36
1.5 Conclusion	38
LA THEORIE DES POSSIBILITES	39
2.1 Introduction: Les données incertaines	40
2.2 Définitions et propriétés des possibilités	43
2.2.1 Liens entre les possibilités et les ensembles flous	43
2.2.2 Liens avec les probabilités	44
2.2.3 Axiomes généraux	45
2.2.4 Hypothèses du monde ouvert et du monde fermé	47
2.2.5 Sens des mesures de possibilités	48
2.2.6 Représentations usuelles	49
2.2.6.1 Information précise et certaine	50
2.2.6.2. Information imprécise	50

2.2.6.3	Information incertaine	51
2.2.6.4	Ignorance absolue.....	52
2.2.7	Construction des degrés de possibilités.....	53
2.2.7.1	Compatibilité entre distributions de probabilités et de possibilités	53
2.2.7.2	Construction de degrés de probabilités à partir d'histogrammes.....	54
2.3	Fusion de données	54
2.3.1	Définitions et opérateurs classiques.....	56
2.3.1.1	Intersection: la fusion conjonctive	56
2.3.1.2	Union: la fusion disjonctive	59
2.3.1.3	Fusion adaptative de Dubois et Prade	60
2.3.1.4	Fusion linéaire.....	61
2.3.2	Fusion adaptative de type logique.....	63
2.3.2.1	Opérateur adaptatif saturé	63
2.3.2.2	Opérateur adaptatif linéaire.....	65
2.3.2.3	Opérateur de type adaptatif hybride	66
2.3.3	Opérateurs de compromis	67
2.3.4	Intégrales floues	69
2.3.5	Fusion basée sur les degrés de confiances.	71
2.3.5.1	Opérateur de priorité	71
2.3.5.2	Fusion linéaire avec coefficients de confiance.....	72
2.3.5.3	Fusion basée sur les transformations d'incertitude.....	72
2.3.6	Conclusion sur les modes de fusion.....	73
2.4	Un nouvel opérateur de fusion.....	74
2.4.1	Objectifs du nouvel opérateur	74
2.4.2	Construction du nouvel opérateur.....	75
2.4.2.1	Structure générale.....	75
2.4.2.2	Conditions limites	75
2.4.2.3	Version π_{Xor} du nouvel opérateur.....	76
2.4.2.4	Version π_{Nand} du nouvel opérateur	77
2.4.2.5	Version π_{Nor} du nouvel opérateur.....	78
2.4.2.6	Comparaisons des opérateurs.....	78
2.4.3	Estimation de la crédibilité accordée au résultat.....	84
2.4.3.1	Composantes de la fiabilité du résultat	85
2.4.3.2	Fiabilité externe.....	86
2.4.3.3	Contradiction interne.....	87
2.4.3.4	Fusion de la fiabilité externe et de la contradiction interne	89
2.4.4	Estimation des crédibilités	90
2.4.4.1	Calcul de la fiabilité dans un cadre probabiliste	90
2.4.4.2	Objectif de l'algorithme d'estimation des crédibilités	91
2.4.4.3	Estimation de la fiabilité interne	92
2.4.4.4	Fusion des fiabilités externes et internes.....	94
2.4.5	Optimisation des degrés de crédibilité	95
2.4.5.1	Construction du problème d'optimisation	95
2.4.5.2	Résolution du problème	97
2.4.5.3	Exemple d'optimisation.....	98
2.4.6	Exemple final: le paradigme du policier contre les voleurs	102
2.5	Conclusion: vers une théorie des possibilités étendue?	105

LE CONTROLEUR FLOU	107
3.1 Introduction: Le succès du contrôleur flou	108
3.2. Structure du contrôleur flou	109
3.2.1 Base de règles.....	109
3.2.1.1 Structure générale.....	109
3.2.1.2 Les termes linguistiques	111
3.2.2 Système de fuzzification	113
3.2.3 Système de défuzzification.....	115
3.2.3.1 Méthode du maximum	115
3.2.3.2 Méthode du centre de gravité.....	116
3.2.3.3 Méthode du centre de gravité pondéré	117
3.2.3.4 Combinaison des méthodes de défuzzification	117
3.2.3.5 Compatibilité entre la fuzzification et la défuzzification	117
3.2.4 Systèmes d'inférences.....	118
3.2.4.1 Structure générale du système d'inférence	118
3.2.4.2 Choix des opérateurs.....	119
3.3 Résultats divers	120
3.3.1 Fonctions élémentaires floues	120
3.3.1.1 Degré de confiance des règles	121
3.3.1.2 Les briques élémentaires en tant qu'outil de construction.....	122
3.3.2 Perturbations dues à des entrées floues.....	123
3.3.3 Liens avec les contrôleurs PID	125
3.3.3.1 Cas d'un contrôleur à une variable	125
3.3.3.2 Cas d'un contrôleur à deux entrées.....	126
3.3.4 Forme analytique exacte.....	126
3.3.5 Simplification de contrôleurs flous.....	128
3.3.5.1 La base de règles existe déjà.....	129
3.3.5.2 Le contrôleur flou n'est pas construit.....	130
3.4 Conclusion: Apport de la logique floue en automatique	132
FONDATEURS DES APPROCHES MULTI-MODELES	135
4.1 Origine des approches multi-modèles	136
4.2 Structure générale des approches multi-modèles.....	137
4.3 Les modèles	141
4.3.1 Classification floue	142
4.3.2 Systèmes d'équations de relations floues	143
4.3.3 Type de modèles fréquents.....	144
4.4 Estimation de la pertinence des modèles.....	145
4.4.1 Formalisme flou	146
4.4.2 Approche géométrique	146
4.4.3 Approche Probabiliste.....	149
4.4.4 Approches des résidus.....	150
4.4.5 Opérations sur les validités finales.....	153
4.5 Fusion des modèles.....	154
4.5.1 Fusion linéaire.....	155

4.5.2 Fusion rencontrée dans les contrôleurs flous	155
4.5.3 Opérateur non linéaire	155
4.5.4 Etape de défuzzyfication	156
4.6 Exemple de contrôleur à validité a priori	156
4.6.1 Le système	156
4.6.2 Les modèles	157
4.6.2.1 Modèles dans le cas d'une configuration à trois modèles	158
4.6.2.2 Modèles dans le cas d'une configuration à 2 modèles	158
4.6.3 Les commandes	158
4.6.4 Calcul des validités a priori	159
4.6.5 Résultats en boucle ouverte avec une configuration à 3 modèles	160
4.6.6 Résultats en boucle fermée avec une configuration à 3 modèles	161
4.6.7 Résultats en boucle fermée avec une configuration à 2 modèles	162
4.6.8 Commande optimale et CMM avec une infinité de modèles	163
4.7 Exemple: Contrôleur à validité estimée par les résidus.....	166
4.7.1 Premier exemple	166
4.7.2 Système perturbé	169
4.8 Conclusion: utilisation des approches multi-modèles.....	172
TECHNIQUES AVANCEES AVEC LE CONTROLEUR MULTI-MODELES	175
5.1 Introduction: Et maintenant, que faire?.....	176
5.2 Bibliothèque de commandes.....	176
5.2.1 Structure d'un contrôleur à plusieurs modèles et commandes	177
5.2.1.1 Observateurs.....	178
5.2.1.2 Validité des lois de commande	178
5.2.1.3 Validité d'une commande élémentaire	179
5.2.2 Exemple de contrôleur à plusieurs commandes	179
5.2.2.1 Notions de la commande à modes glissants	180
5.2.2.2 Les données.....	181
5.2.2.3 Le contrôleur à plusieurs commandes.....	181
5.2.2.4 Résultat classique.....	182
5.2.2.5. Résultats avec deux modèles bien espacés.....	183
5.2.2.6. Résultats avec deux modèles mieux choisis.	184
5.3 Adaptation de l'estimation des validités.....	185
5.3.1 Ajustement des validités fournies par les résidus	185
5.3.1.1 Retour sur l'approche classique des résidus	185
5.3.1.2 Stratégies d'adaptation.....	187
5.3.1.3 Exemple avec le second système perturbé	188
5.3.2 Optimisation des validités avec l'étape de renforcement	192
5.3.2.1 Formulation du problème.....	192
5.3.2.2 Modélisation d'une série temporelle	193
5.3.4 Complexité des représentations multi-modèles	198
5.4 Perspectives.....	199
5.4.1 Très grande bibliothèque.....	199
5.4.2 Application en surveillance.....	200
5.5 Conclusion.....	202

BIBLIOTHEQUE DYNAMIQUE DE MODELES	203
6.1 Pourquoi une bibliothèque de modèles?	204
6.2 Structure d'un contrôleur avec un seul modèle identifié.....	206
6.3 Identification de systèmes non linéaires.....	207
6.3.1 Modèles pour l'identification	208
6.3.2 Méthodes adaptatives	208
6.3.3 Méthodes de segmentation.....	208
6.3.4 Modèles à coefficients aléatoires	209
6.3.5 Modèles évolutifs	209
6.3.6 Choix de la méthode d'identification	210
6.4 Vers la Multi-Identification.....	210
6.4.3.1 Identification pure du système complexe.....	211
6.4.3.2 Bouclage de l'identification du système complexe.	212
6.4.3.3 Bouclage de l'identification, cas de la poursuite.	215
6.4.4 La multi-identification	217
6.5 Conclusion.....	217
CONCLUSION GENERALE.....	219
BIBLIOGRAPHIE	223
A LES ALGORITHMES GENETIQUES	229
B IDENTIFICATION DE MODELES DISCRETS	233
B.1 Algorithmes d'identification.....	234
B.1.2 Algorithme des moindres carrés à facteur d'oubli	234
B.1.3 Algorithmes des moindres carrés à fenêtre glissante.....	235
B.2.2.3 Algorithmes avec excitation permanente	236
B.2. Conversions modèle continu modèle discret	237
B.3 Filtres du modèle identifié	238
B.3.1 Structure des filtres.....	238
B.3.2 Evaluation du gain α	238
B.3.3 Filtre proche de la surveillance.....	239
B.3.4 Filtre linéaire	239
B.4 Modélisation avec les bases de fonctions	240
B.4.1 Bases de fonctions	240
B.4.2 Signal dans la base.....	240
B.4.2.1 Première simulation.....	240
B.4.2.2 Seconde simulation, base petite.....	242
B.4.2.3 Seconde simulation, base trop grande	244
B.4.3 Signal en dehors de la base.....	246
B.4.3.1 Signal un peu hors de la base	246
B.4.3.2 Signal moyennement hors de la base.....	251
B.4.3.3 Signal totalement hors de la base	253
B.4.4 Conclusion sur l'identification avec base de fonctions.....	257

INTRODUCTION GENERALE

Principe du Gestalt ou Théorie de la Forme

Théorie moderne selon laquelle les propriétés d'un phénomène psychique ou d'un être vivant ne résultent pas de la simple addition des propriétés de ses éléments, mais de l'ensemble des relations entre ces éléments

Petit Robert.

Jusqu'à récemment encore, le fossé s'élargissait entre d'une part la réflexion humaine, et d'autre part les sciences dites exactes.

En effet la première utilise sans peine des concepts mal identifiés, imprécis, vagues. Par exemple un homme dit d'une eau qu'elle est chaude ou froide, voire tiède, sans pour autant définir précisément sa température. De même une voiture peut être caractérisée comme se déplaçant à vive allure ou au contraire lentement, sans pour autant avoir sa vitesse définie au km/h près. Il serait possible de citer un nombre presque illimité de variables ou de situations dans lesquelles l'homme arrive à analyser et à agir avec à sa disposition uniquement des variables non clairement quantifiées (les deux exemples cités), incertaines (la fiabilité des informations est douteuse), ou même incomplètes.

Cette faculté apparaît inhérente au cerveau humain, comme l'ont montré récemment des chercheurs étudiant des personnes atteintes de lésions spécifiques du cerveau. En effet ces personnes, à la suite d'une maladie ou d'un accident, se révélèrent incapables de faire des calculs que se soit à la main ou de tête. Par exemple elles étaient dans l'incapacité de donner le résultat de $5+12$. Par contre elles pouvaient affirmer avec force que 73×17 était totalement différent de 32105.

C'est d'une toute autre logique que procèdent les sciences actuelles. Celles-ci ont permis la réalisation d'entreprises grandioses, comme par exemple à l'homme d'aller sur la Lune. Mais pour cela, il aura fallu -et il faudra toujours- fournir aux règles à calcul, ordinateurs et autres calculatrices, des quantités énormes de chiffres précis qui sont ensuite traités dans tel ou tel algorithme pour résoudre telle ou telle équation.

Les sciences exactes exigent donc des estimations très précises des variables mesurées, qui sont alors considérées comme sûres, et si une variable cruciale n'a pas été évaluée, l'algorithme s'arrête. Si d'aventure les variables sont parasitées, les bruits sont alors supposés appartenir à telle ou telle classe, et des méthodes statistiques élaborées permettent de les éliminer. Mais si ces parasites ou ces imperfections des variables ne satisfont pas les hypothèses envisagées, les algorithmes, une fois de plus, s'arrêtent.

Ce fossé, entre Sciences et Hommes, aurait pu s'élargir encore. Ainsi certaines études sociologiques ont montré qu'une fraction de plus en plus importante des populations de notre monde considérait les techniques les plus avancées de notre ère comme de la magie pure et simple.

Cependant l'apparition de la logique floue a permis la prise en compte de l'incertain de manière rigoureuse et scientifique. Le succès de la logique floue dans la recherche scientifique est phénoménal, et peu de nouvelles théories ont autant révolutionné la recherche scientifique dans ses méthodes mêmes. Agée de presque trente ans, le nombre de publications ayant trait à son domaine n'a cessé d'augmenter, et ses applications dans l'industrie, lancées par les Japonais, connurent le même succès.

Ceci peut s'expliquer simplement par l'atout majeur de la logique floue, c'est à dire celui de pouvoir créer des fonctions complexes avec des outils simples et surtout compréhensibles par le concepteur.

A l'origine les applications de la logique floue furent essentiellement basées sur les connaissances de praticiens experts, qui étaient alors représentées puis programmées grâce à la logique floue.

En effet si le cerveau est encore considéré comme une boîte noire, ses interactions avec le monde sont habituellement représentées de manière séquentielle. Par exemple un savoir d'expert est transcrit sous la forme d'une base de règles liant des observations sur des variables à des conclusions, selon le célèbre aphorisme

"Si observation est Alors conclusion est".

Or la logique floue permet non seulement de représenter et de traiter les termes imprécis utilisés par l'expert dans sa description des faits et conclusions, mais aussi de fusionner l'ensemble des règles pour obtenir une seule conclusion concernant une certaine variable de sortie d'un système, une commande, une action dans un système d'aide à la décision, ou encore tout autre élément désiré, de manière à pouvoir "mimer" la réflexion humaine.

Une fonction complexe a alors été réalisée par l'intermédiaire d'une série de règles ou de fonctions élémentaires très simples et lisibles par un être humain.

Cette facilité apparut de prime abord, et masqua l'autre révolution conceptuelle apportée par cette représentation: la représentation de phénomènes ou de systèmes complexes par une série de connaissance partielles.

Dans l'exemple de la base de règles fournie par un humain, chaque conclusion est une connaissance partielle, et elle n'est valide que sous certaines circonstances, représentées par les conditions de la règle.

Par contraste, en choisissant l'exemple de l'automatique, lorsqu'un système doit être commandé par une commande optimale, il faut d'abord un modèle exact dudit système. Ce modèle doit refléter de la manière la plus précise possible le système considéré, être valide pour toutes les situations possibles, prendre en compte tous les comportements... Ensuite une loi de commande sera calculée. Une distinction entre modèle de connaissance et modèle de représentation fut introduite pour résoudre la complexité des modèles obtenus. Les modèles de connaissances sont fournis par l'application des lois connues de la physique, et donnent en général des modèles extrêmement complexes. Au contraire les modèles de représentation ont pour objectif de simplifier la représentation, pour l'étude des systèmes ou le calcul de commandes. Cependant même cette démarche rigoureuse est de plus en plus mise en défaut, alors que les sciences affrontent des systèmes ou des phénomènes dont la complexité croît sans cesse. Il est de plus en plus difficile d'obtenir des modèles précis et corrects représentant les systèmes dans toutes leurs particularités. Il arrive même, ultime ironie, qu'un modèle, de représentation, global et précis, puisse être défini, mais se révèle alors beaucoup trop complexe pour être utilisable.

C'est alors qu'une approche *multi-modèle* peut être choisie. A l'image de la base de règle de la logique floue, l'approche multi-modèle n'essaie pas de représenter dans leur ensemble des phénomènes complexes, mais bien au contraire a pour objectif de donner un ensemble d'aperçus, ou de représentations partielles, valides et applicables uniquement sous certaines conditions bien spécifiques. Chaque vue partielle, ou modèle, est pratiquement toujours faux, sauf dans quelques cas particuliers, et ceci pour la totalité des vues partielles. Si l'ensemble a été construit de telle sorte qu'au moins une vue partielle corresponde approximativement au système pour un environnement donné, la fusion des différentes représentations partielles, par

un mécanisme à déterminer, donne une vue globale, générale, exacte et précise du système réel. Ceci avec une série de modèle qui peuvent être beaucoup plus simples que le modèle théorique complet, simplicité qui ultérieurement facilite l'étude du système en vue de la détermination d'une commande, d'une aide à la décision ou d'un diagnostic, suivant le champ d'application considéré.

L'approche multi-modèle apparaît donc pleine d'avenir pour représenter et gérer les systèmes très complexes, et l'objectif de cette thèse est de montrer sa pertinence, notamment dans le cadre de l'automatique, et de faire un état des travaux les plus récents la concernant. Bien que l'approche développée ici commence à s'abstraire de ses racines floues, d'autres approches multi-modèles sont encore basées sur elles.

Aussi dans les premières parties de ce mémoire seront exposés quelques principes de la logique floue. La première partie est relative aux définitions élémentaires de la théorie des ensembles flous, la seconde partie concerne la théorie des possibilités, issue de la logique floue, car le schéma de contrôleur multi-modèle présenté finalement utilise des principes de cette théorie.

La troisième partie présente les principes du contrôleur flou classique, qui amène naturellement à l'idée du contrôleur multi-modèles.

La quatrième partie expose les premiers principes de ces contrôleurs, principes qui sont complétés par des raffinements présentés dans la cinquième partie. La dernière partie propose un schéma de contrôleur multi-modèle qui s'adapte aux évolutions des systèmes.

CHAPITRE PREMIER

DES ENSEMBLES FLOUS A LA LOGIQUE FLOUE

1.1 Introduction: Pourquoi les ensembles flous

La théorie classique des ensembles est essentiellement binaire: un élément appartient ou non à un ensemble, et il n'y a pas d'autre possibilité. C'est pourquoi il était relativement difficile de faire cadrer le raisonnement humain avec cette logique. En effet les objets manipulés par les humains sont beaucoup plus nuancés et imprécis, ce qui ne pouvait être traduit qu'imparfaitement avec cette logique.

Zadeh proposa en 1965 une nouvelle approche [Zadeh, 1965] avec laquelle il introduisait les ensembles flous. Comme leur nom l'indique, les ensembles flous peuvent être caractérisés de manière beaucoup moins tranchée que les ensembles classiques, de manière beaucoup plus vague...

Ils permirent de mimer les données traitées par les humains, et les algorithmes mis en jeu par eux, pour aboutir notamment à la logique floue. En 1974 Mamdani définit une méthode de commande basée sur la logique floue, appelée contrôleur flou. En 1980 les premières applications industrielles virent le jour au Japon, et depuis lors le nombre de publications ayant trait à la logique floue ne cesse de croître.

L'objectif de ce court chapitre est de montrer les liens des ensembles flous avec les ensembles classiques. Les opérations élémentaires sur ces derniers, comme l'union ou l'intersection, seront alors généralisées. Quelques exemples d'opérations plus complexes sont donnés afin de montrer la richesse des ensembles flous. Finalement les principes de la logique floue sont abordés avec la généralisation du syllogisme.

1.2 Définition d'un ensemble flou

'Mais à la fin, qu'est un ensemble flou?' pourrait se demander un non spécialiste, lassé de voir cette expression lancée à tout va. Dans cette partie la définition des ensembles flous et les liens avec la théorie classique seront donnés.

1.2.1 Origine des ensembles flous

Soit un ensemble classique $A \subset X$. L'appartenance à celui-ci est binaire: un élément $x \in X$ appartient ou non à A . Dans la théorie classique des ensembles, sont associées à chaque ensemble A les fonctions I_A appelées fonctions caractéristiques:

$$I_A: X \rightarrow \{0,1\}$$

Les ensembles flous sont aussi caractérisés par une fonction caractéristique, mais dans leur cas, elle s'appelle fonction d'appartenance μ_A . L'idée "géniale" de Zadeh [Zadeh, 1965] a été de ne pas contraindre cette fonction à prendre ses valeurs dans $\{0, 1\}$, mais dans $[0,1]$:

$$\mu_A: X \rightarrow [0,1]$$

Ainsi cette fonction peut prendre d'autres valeurs que 0 ou 1, par exemple 0.5: $\mu_A(x)=0.5$ signifie simplement que x appartient à A avec un degré égal à 0.5.

L'exemple qui suit va montrer la facilité qu'ont les ensembles flous de décrire des concepts imprécis.

Souvent pour régler la température de l'eau de son bain, l'homme trempe sa main dans l'eau. Cet outil universel a beaucoup de fonctions, mais il n'est pas un capteur de température

absolue très précis (bien que certaines études montrent qu'il est capable de détecter des variations de 1° Celsius). De toute façon un homme ne dit pas d'une eau qu'elle est à 20,6°C, mais plutôt il utilise des labels pour la caractériser, et notamment il dit d'une eau qu'elle est froide, tiède, chaude ou encore bouillante. Une eau à 25°C ne lui apparaîtra plus tout à fait tiède, mais presque chaude, et une eau à 26°C un peu moins tiède et un peu plus chaude.

La théorie des ensembles classique ne permet pas de modéliser ces sensations, et notamment la nuance introduite entre 25°C et 26°C entre tiède et chaud: soit une eau à 25°C est tiède, soit elle est chaude. Il en est de même pour un eau à 26°C, et toute nuance disparaît.

Il en est tout autrement avec les ensembles flous. La figure 1.1 montre une modélisation possible des labels humains, basée sur des fonctions d'appartenance triangulaires ou trapézoïdales.

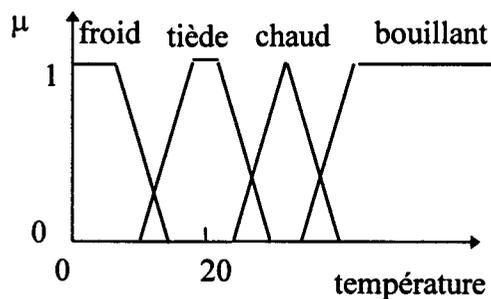


Fig. 1.1. Ensembles flous caractérisant la température de l'eau du bain.

En général, les fonctions d'appartenance utilisées sont de forme triangulaire, trapézoïdale, gaussienne, en forme de cloche...

En conclusion la seule différence, minime en fait, entre les ensembles classiques, et les ensembles flous, porte sur la fonction caractéristique, qui de valeurs dans $\{0,1\}$, passe à des valeurs dans $[0,1]$.

Pourtant les conséquences, comme le montreront les pages qui suivent, auront été très importantes.

1.2.2 Lien avec la théorie classique: les coupures α

Aussitôt créés, les théoriciens ont voulu relier les ensembles flous aux ensembles classiques. Ce lien est réalisé grâce aux α -coupes.

Soit un ensemble flou A donné par sa fonction d'appartenance μ_A .

Etant donné un scalaire $\alpha \in]0,1]$ une coupure α est définie comme un ensemble classique A_α donné par la relation suivante:

$$x \in A_\alpha \Leftrightarrow \mu_A(x) \geq \alpha \quad (1.1)$$

Ainsi un élément appartient à une coupure A_α si et seulement si sa valeur d'appartenance à l'ensemble flou A est au moins égale au niveau de coupe α .

A l'inverse, étant donné une série d' α -coupes, la relation suivante permet de retrouver l'ensemble flou originel:

$$\mu_A(x) = \max_{\alpha} \alpha \times \text{Booléen}(x \in A_{\alpha}) \quad (1.2)$$

(Booléen étant une fonction binaire suivant la valeur de vérité de son argument).

Soit l'ensemble flou caractérisé par sa fonction d'appartenance tracée sur la figure 1.2:

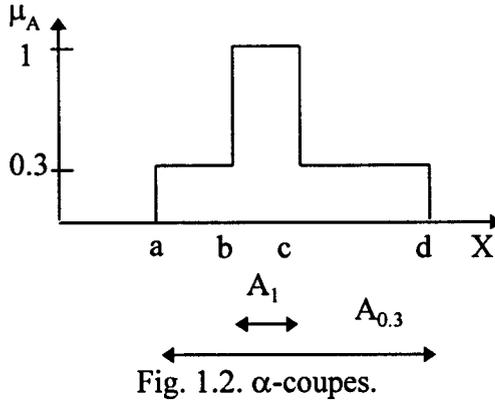


Fig. 1.2. α -coupes.

L'ensemble de référence X est un compact. Toutes les valeurs x comprises entre b et c appartiennent en totalité à l'ensemble A , et les x compris entre a et b puis entre c et d n'appartiennent à A qu'un à degré égal à 0.3 . Toute autre valeur de x n'appartient pas du tout à A .

Dans ce cas simple, l'ensemble minimal d' α -coupes permettant de reconstruire l'ensemble flou original se réduit à 2: A_1 et $A_{0.3}$. Par exemple $A_1 = [b, c]$, car $\forall x \in [b, c], \mu_A(x) = 1$.

Cet exemple permet d'introduire deux quantités qui sont souvent utilisées pour caractériser un ensemble flou: son cœur et son support.

Le cœur est l'ensemble classique des éléments pour lesquels la fonction d'appartenance vaut 1.

$$\text{Cœur}(A) = A_1 \quad (1.3)$$

(A_1 étant la coupure de niveau 1)

Le support est l'ensemble classique des éléments tels que la fonction d'appartenance ait une valeur différente de 0, c'est à dire ceux qui appartiennent, même très faiblement, à l'ensemble flou.

Grâce aux coupures α , il est possible de caractériser cet ensemble de la manière suivante:

$$\text{Support}(A) = A_{\min_x(\mu_A(x); \mu_A(x) > 0)} \quad (1.4)$$

Si le cœur d'un ensemble flou n'est pas l'ensemble vide, il est alors dit normalisé.

Les coupures α sont utilisées comme liens avec la théorie des ensembles classiques, et servent dans des démonstrations de certains théorèmes ou propriétés.

1.3 Négation, Union, Intersection

La définition des ensembles flous ayant été donné, il s'agit d'aborder les opérations élémentaires qui s'appliquent à eux.

1.3.1 Définitions Classiques

Les exemples seront basés sur les deux ensembles flous suivants, caractérisés par leur fonctions d'appartenances:

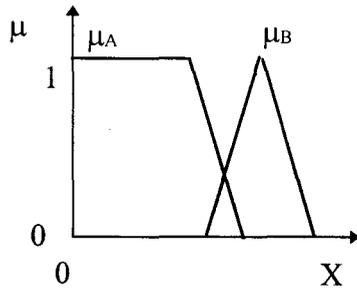


Fig. 1.3. Ensembles A et B de l'exemple.

1.3.1.1 Complément

Le complément d'un ensemble flou caractérisé par $\mu(x)$ est donné par:

$$\mu^c(x) = 1 - \mu(x) \tag{1.5}$$

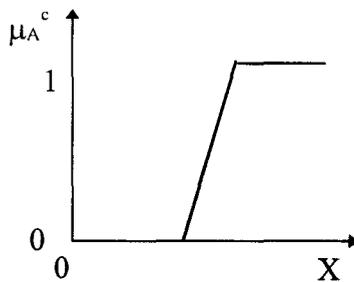


Fig. 1.4. A^c

Cette définition est parfaitement cohérente avec celle de la théorie classique. Soit un ensemble C associé à sa fonction caractéristique I_C . Si un élément x appartient à C, $I_C(x)=1$, mais x n'appartient pas au complément de C, C^c , alors $I_C^c(x)=0$, soit $1-I_C(x)$. La réciproque tient pour un x' n'appartenant pas à C.

1.3.1.2 Intersection

L'intersection de deux ensembles flous A et B est donnée par:

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \tag{1.6}$$

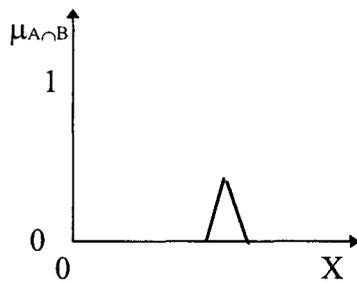


Fig. 1.5. $A \cap B$

Il est possible de vérifier si cette définition de l'intersection floue est compatible avec l'intersection classique: Pour qu'un élément x appartienne à l'intersection de deux ensembles C et D, il faut et il suffit qu'il appartienne aux deux à la fois. Donc $I_C(x)=1$ et $I_D(x)=1$ implique que x soit dans l'intersection. Au contraire pour un x' différent, dès que $I_C(x')=0$ ou $I_D(x')=0$, x' n'appartient plus à l'intersection. Donc l'opérateur min s'applique aussi dans la théorie classique.

Il est possible de vérifier si cette définition de l'intersection floue est compatible avec l'intersection classique: Pour qu'un élément x appartienne à l'intersection de deux ensembles C et D , il faut et il suffit qu'il appartienne aux deux à la fois. Donc $I_C(x)=1$ et $I_D(x)=1$ implique que x soit dans l'intersection. Au contraire pour un x' différent, dès que $I_C(x')=0$ ou $I_D(x')=0$, x' n'appartient plus à l'intersection. Donc l'opérateur min s'applique aussi dans la théorie classique.

1.3.1.3 Union

L'union entre deux ensembles flous A et B est donnée par la relation:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (1.7)$$

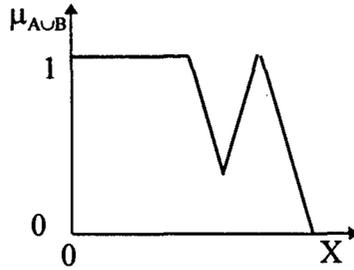


Fig. 1.6. $A \cup B$

L'opérateur max, comme l'opérateur min pour l'intersection, est parfaitement cohérent avec la théorie classique.

1.3.2 Généralisation des opérateurs avec les t -normes et t -conormes

Les opérateurs précédemment définis pour généraliser aux ensembles flous les règles classiques de négation, union et intersection, n'ont pu être vérifiés que dans le cas limite où les ensembles flous considérés étaient en fait des ensembles classiques (les fonctions d'appartenance ne prenant comme valeurs que 1 ou 0).

Il est légitime de se demander si d'autres opérateurs n'auraient pas pu être choisis, avec bien évidemment les mêmes comportements limites dans le cas d'ensembles classiques.

Pour cela certaines propriétés ont été imposées aux éventuels opérateurs, et il apparut que, outre les opérateurs min, max et le complément à 1, d'autres opérateurs pouvaient être définis pour jouer le rôle, qui d'une intersection, qui d'une union, qui d'un complément.

Les familles de ces opérateurs généralisent les opérateurs max, min et complément à 1, et sont présentés ci-dessous. Ils appartiennent aux familles des t -normes, respectivement des t -conormes, dans le cas de l'intersection, respectivement dans le cas de l'union [Schweize, 1967; Weber 1983].

1.3.2.1 Les t -normes et t -conormes

-Une fonction $\cap: [0,1] \times [0,1] \rightarrow [0,1]$ est une t -norme si:

\cap est commutative, associative, non décroissante, et a 1 comme élément neutre

Toute t -norme peut jouer le rôle d'une intersection en logique floue.

-Une fonction $\cup: [0,1] \times [0,1] \rightarrow [0,1]$ est une t -conorme si:

\cup est commutative, associative, non décroissante et a 0 comme élément neutre
Toute t-conorme peut jouer le rôle d'une union.

-Une fonction $n:[0,1] \rightarrow [0,1]$ est dite négation (ou jouant le rôle de complément) si:
 $n(0)=1, n(1)=0$, et n est non croissante.

-Etant donné un opérateur de complément, deux opérateurs d'intersection \cap et d'union \cup sont dits duaux s'ils respectent la relation:

$$\cap (n(x), n(y)) = n(\cup(x, y)) \tag{1.8}$$

Afin d'avoir des comportements symétriques et compréhensibles suivant la logique classique, il est préférable d'utiliser des opérateurs duaux.

Avec les t-normes et t-conormes, il est maintenant possible de choisir d'autres opérateurs que le min et le max. Dans la suite, les termes t-normes, respectivement t-conormes, sont indifféremment utilisés pour l'intersection, respectivement l'union, et inversement.

Par exemple l'opérateur produit est souvent utilisé en logique floue pour généraliser l'intersection.

Il est possible de vérifier sa cohérence avec la théorie classique: pour qu'un élément x appartienne à l'intersection de deux ensembles C et D , il faut et il suffit qu'il appartienne aux deux à la fois. Donc $I_C(x)=1$ et $I_D(x)=1$ implique que x soit dans l'intersection, ce qui est vérifié avec $I_C(x) \times I_D(x)=1$

Au contraire pour un x' tel que $I_C(x')=0$ ou $I_D(x')=0$, x' n'appartient pas à l'intersection. Or $I_C(x') \times I_D(x')=0$ dans un de ces cas-là. En définitive, l'opérateur produit est parfaitement compatible avec la théorie des ensembles classiques.

1.3.2.2 Exemples de t-normes et t-conormes

Le tableau 1.1 donne quelques exemples d'opérateurs duaux jouant le rôle d'intersection et d'union.

t-normes	t-conormes duales	noms
$\cap_m = \min(x, y)$	$\cup_m = \max(x, y)$	Zadeh
$\cap_x = xy$	$\cup_x = x + y - xy$	Probabiliste
$\cap_y = 1 - \min(1, ((1-x)^p + (1-y)^p)^{1/p})$	$\cup_y = \min(1, (x^p + y^p)^{1/p})$	Yager, $p \geq 0$
$\cap_D = \begin{cases} x & \text{si } y = 1 \\ y & \text{si } x = 1 \\ 0 & \text{autrement} \end{cases}$	$\cup_D = \begin{cases} x & \text{si } y = 0 \\ y & \text{si } x = 0 \\ 1 & \text{autrement} \end{cases}$	Discontinue
$\cap_L = \max(x + y - 1, 0)$	$\cup_L = \min(x + y, 1)$	Lukasiewicz

Tableau 1.1. Quelques t-normes et t-conormes.

La famille des opérateurs de Zadeh est la plus utilisée. Les opérateurs probabilistes sont aussi beaucoup utilisés, car ils ont des comportements beaucoup plus nuancés que les opérateurs de Zadeh. En effet soit le calcul de $\cap(x, y)$. Avec les opérateurs de Zadeh (en l'occurrence un min) le résultat est identique quelque soit la valeur de x tant qu'elle reste supérieure à celle de

y, avec les opérateurs probabilistes (en l'occurrence le produit), le résultat varie continûment en fonction de x.

Les opérateurs de Lukasiewicz sont aussi intéressants, car hormis leurs bornes de variations (0 pour l'intersection et 1 pour l'union), leurs variations sont linéaires. Il existe un grand nombre de t-normes et t-conormes, dont quelques familles qui dépendent d'un paramètre, et permettent de retrouver certains opérateurs remarquables, comme les opérateurs de Yager.

Etant donnés une intersection \cap et une union \cup , la relation d'encadrement suivante est toujours vérifiée:

$$\cap_D \leq \cap \leq \cap_m \leq \cup_m \leq \cup \leq \cup_D \quad (1.9)$$

Toute union est comprise entre l'union de Zadeh, la plus petite, et l'union discontinue, la plus grande. La plus petite des intersections est celle discontinue, et l'intersection la plus grande est celle de Zadeh. Les figures 1.7 à 1.10 montrent les résultats en 3D d'intersections classiques.

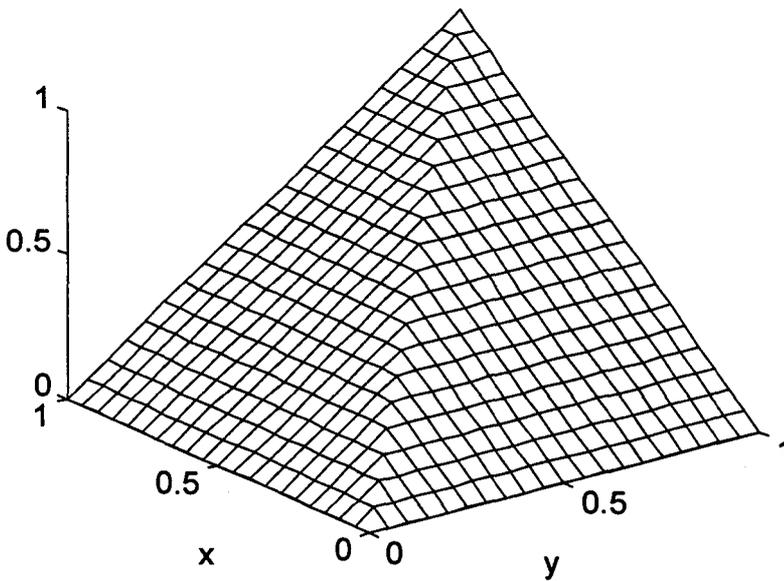


Fig. 1.7. $\cap = \min(x,y)$

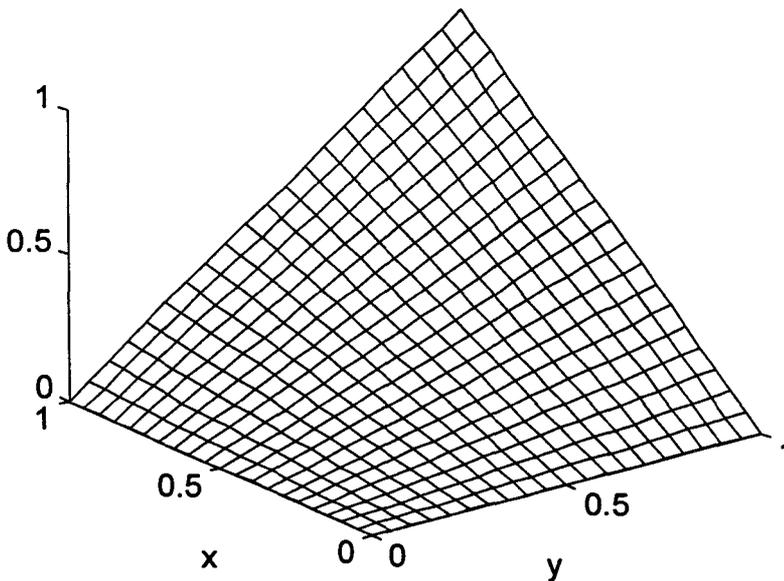


Fig. 1.8. $\cap = xy$

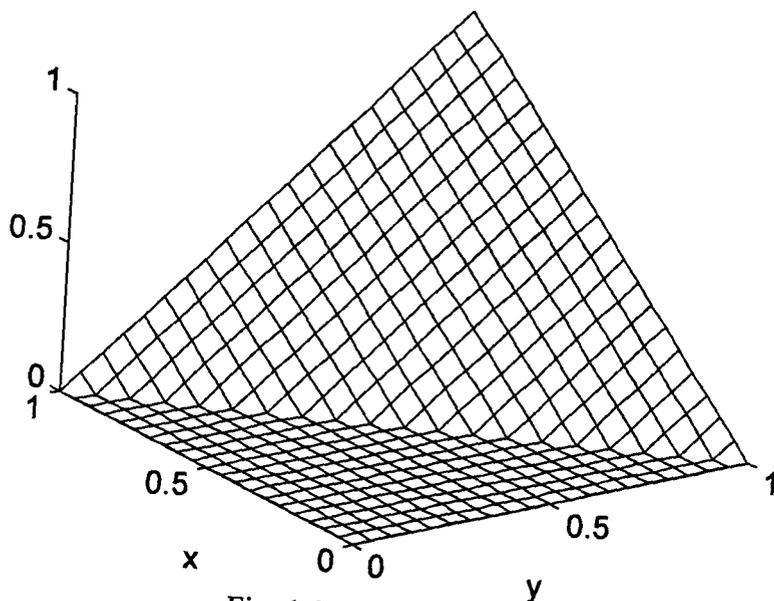


Fig. 1.9. $\mu = \max(x+y-1, 0)$

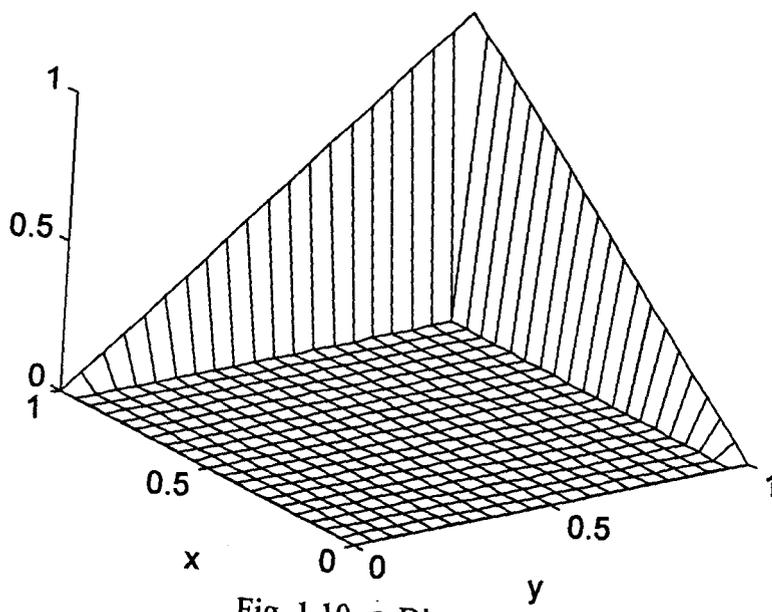


Fig. 1.10. μ Discontinue

Les figures 1.11 à 1.14 montrent les résultats de l'union.

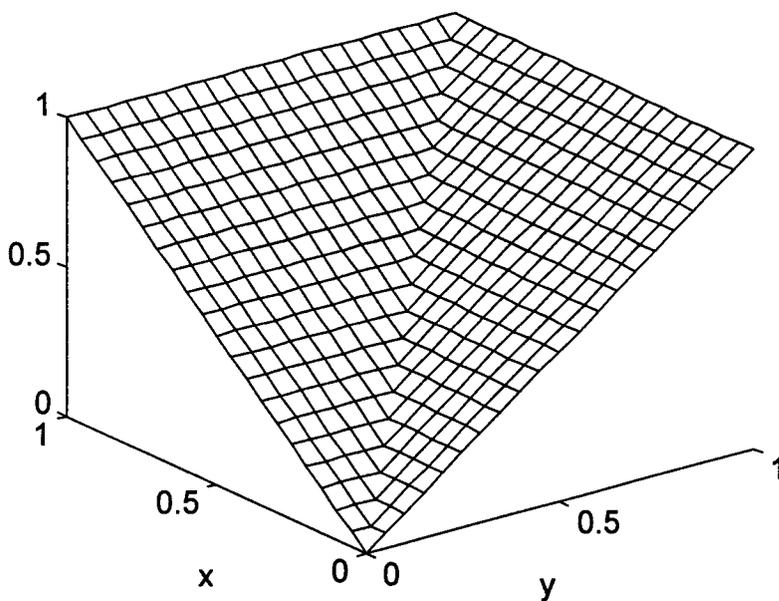


Fig. 1.11. $\cup = \max(x,y)$

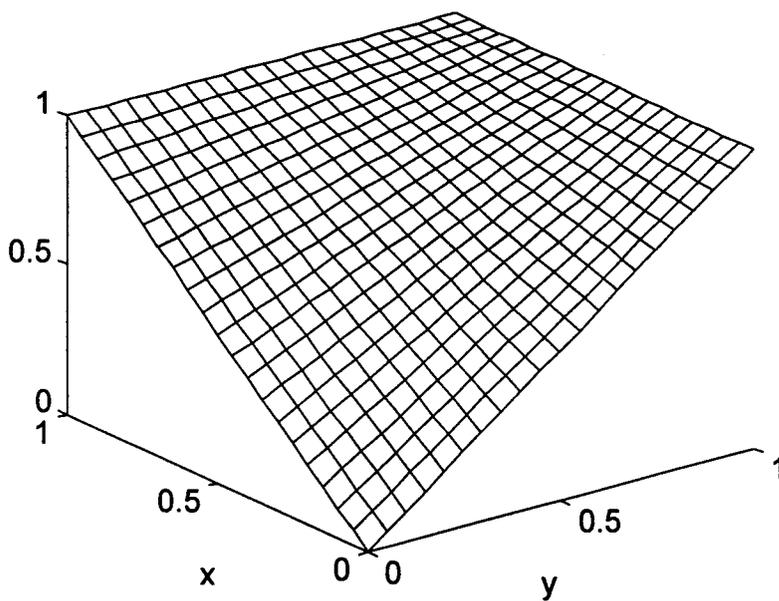


Fig. 1.12. $\cup = x+y-xy$

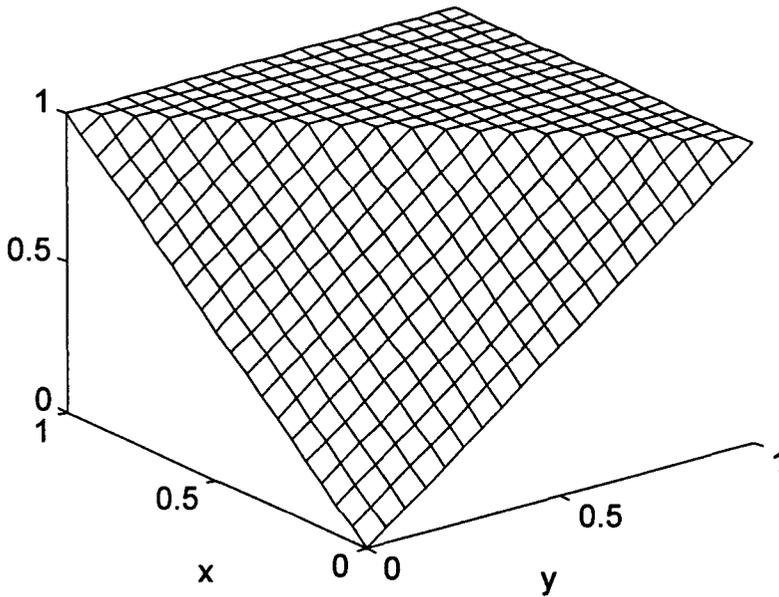


Fig. 1.13. $\cup = \min(x+y, 1)$

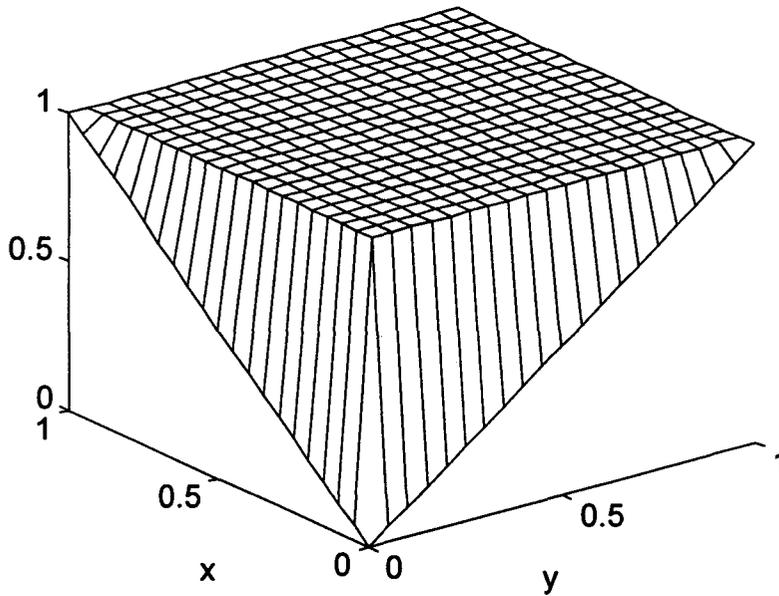


Fig. 1.14. \cup Discontinue

1.3.2.3 Origine des t-normes et t-conormes

Il existe un grand nombre de t-normes et de t-conormes dans la littérature, et elles ont parfois des types très différents (comme par exemple la famille probabiliste et la famille de Lukaziewicz). Dans ce paragraphe est donné un moyen très simple de les construire.

Pour cela il faut revenir sur les définitions des t-normes et t-conormes. En effet celles données précédemment étaient les plus faibles qui soient.

Une t-norme \cap est dite Archimédienne si:

- \cap est continue
- $\cap(x, x) < x$ pour $x \in]0, 1[$

Une t-norme Archimédienne est dite stricte si elle est strictement croissante sur l'ouvert $]0, 1[$.

Une t-conorme \cup est dite Archimédienne si:

- \cup est continue

- $\cup(x,x) > x$ pour $x \in]0,1[$.

Elle est dite stricte si elle est strictement croissante sur $]0,1[$.

Il est possible de construire toutes les t-normes et t-conormes strictes par l'intermédiaire de fonctions appelées générateurs additifs.

Soit une fonction $f: [0,1] \rightarrow [0,\infty[$ avec $f(1)=0$, décroissante et continue.

D'abord est définie la pseudo inverse f^{-1*} de f par:

$$f^{-1*}(y) = \begin{cases} f^{-1}(y) & \text{si } y \in [0, f(0)] \\ 0 & \text{si } y \in [f(0), \infty] \end{cases} \quad (1.10)$$

Alors

$$f^{-1*}(f(x)+f(y)) \quad (1.11)$$

est une t-norme.

De même soit une fonction $g: [0,1] \rightarrow [0,\infty[$, avec $g(0)=0$, croissante et continue.

La pseudo inverse g^{-1*} de g est définie alors par:

$$g^{-1*}(y) = \begin{cases} g^{-1}(y) & \text{si } y \in [0, g(1)] \\ 0 & \text{si } y \in [g(1), \infty] \end{cases} \quad (1.12)$$

Alors

$$g^{-1*}(g(x)+g(y)) \quad (1.13)$$

est une t-conorme.

Par exemple la famille probabiliste a les générateurs suivants:

-pour l'intersection $\cap(x,y)=xy$, f est définie par $f(z)=-\ln(z)$, avec $f^{-1}(z)=\exp(-z)$

-pour l'union $\cup(x,y)=x+y-xy$, g est définie par $g(z)=-\ln(1-z)$ avec $g^{-1}(z)=1-\exp(-z)$.

1.3.2.4 Compatibilité avec la théorie classique

Il existe un très grand nombre de formes d'union et d'intersection avec les t-normes et t-conormes, et il est possible de s'interroger sur le choix de tel ou tel opérateur. Si certaines propriétés sont souhaitées, et qu'un opérateur les respecte, alors il peut être choisi. Ainsi la sensibilité de la famille probabiliste peut être un facteur désiré.

Les opérateurs min et max sont respectivement les plus grands et les plus petits des intersections et unions. Mais ils ont aussi une autre propriété remarquable.

En effet si toutes les opérations de fusion, c'est à dire d'intersection ou d'union, sont réalisées par la théorie des ensembles flous, il est possible d'en vérifier les résultats dans la théorie classique au moyen des coupures α . Pour cela les ensembles flous à traiter sont traduits en séries d'ensembles classiques par le moyen de la relation (1.1), puis fusionnés normalement, en donnant autant de résultats qu'il y avait de coupures α .

Ensuite ces résultats sont utilisés pour reconstruire un ensemble flou par la relation (1.2).

Cet ensemble flou est alors comparé avec le résultat de la fusion obtenue directement dans la théorie des ensembles flous.

Il apparaît alors que les résultats ne sont jamais identiques, sauf si les opérateurs sont exclusivement ceux de Zadeh [Bezdek, 1986].

C'est pourquoi les opérateurs min et max peuvent être considérés comme les vraies généralisations des opérateurs de la théorie classique. Les autres opérateurs peuvent être utilisés, mais il ne faut plus alors chercher à interpréter les résultats en termes de théorie des ensembles classique.

1.3.3 Opérateurs étranges

Les ensembles flous étant caractérisés par des degrés d'appartenance compris entre 0 et 1, et non plus seulement égaux à 0 ou 1, des opérateurs plus complexes ont pu être définis, alors qu'ils n'avaient pas d'équivalents en logique classique. Quelques exemples sont donnés ici.

1.3.3.1 Opérateurs de maximisation et de minimisation

Soit f un générateur additif d'une t -norme.

Un opérateur de maximisation τ est défini par:

$$\tau(x,y) = \begin{cases} 1 & \text{si } x \geq y \\ f^{-1}(f(x) - f(y)) & \text{si } x < y \end{cases} \quad (1.14)$$

Etant donné un générateur additif g d'une t -conorme, un opérateur de minimisation σ est défini par:

$$\sigma(x,y) = \begin{cases} 0 & \text{si } x \leq y \\ g^{-1}(g(x) - g(y)) & \text{si } x \geq y \end{cases} \quad (1.15)$$

Ces deux opérateurs sont utilisés dans la résolution de systèmes d'équations floues.

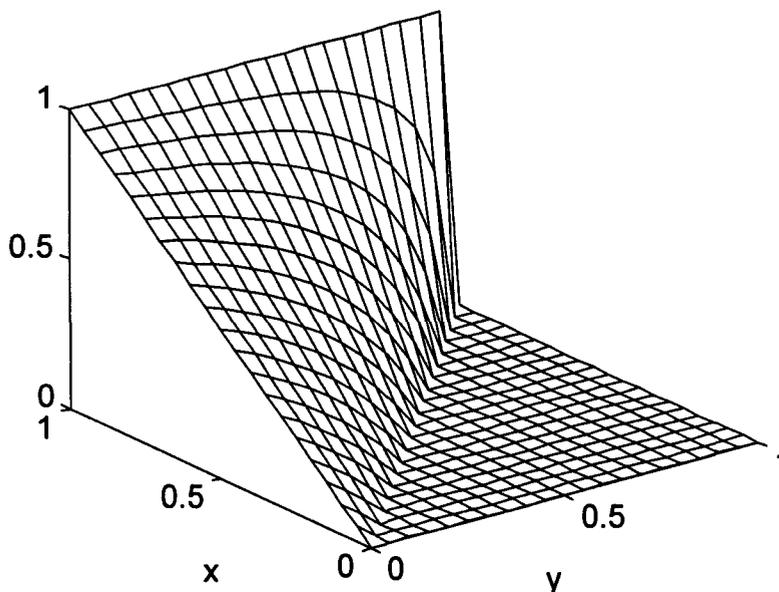


Fig. 1.15. $\sigma(x,y)$ pour le générateur g de l'union probabiliste ($\cup = x+y-xy$)

1.3.3.2 Opérateurs mi-intersection mi-union

Grâce à des coefficients, il est possible de définir des opérateurs ayant des comportements à mi-chemin de l'union et de l'intersection. Dans la théorie classique, cela n'était pas possible. Soit par exemple un hypothétique opérateur 0.5 union +0.5 intersection (en travaillant sur les fonctions caractéristiques). Si cet opérateur avait été appliqué à un élément n'appartenant pas à l'intersection des ensembles à fusionner, le résultat aurait été une valeur caractéristique égale à 0.5, ce qui n'avait aucun sens dans la théorie classique.

Par contre un tel opérateur aurait été parfaitement valable dans la théorie des ensembles flous, et il a d'ailleurs été défini.

Tout d'abord les opérateurs flous $\Phi(x,y)$ sont classés suivant leur comportement [Bloch 1995].

- si $\Phi(x,y) \leq \min(x,y)$ alors Φ est dit conjonctif
- si $\Phi(x,y) \geq \max(x,y)$ alors Φ est dit disjonctif
- si $\min(x,y) < \Phi(x,y) < \max(x,y)$ alors Φ est dit de compromis

Il existe des opérateurs dont le comportement est statique: toujours conjonctif par exemple. Ainsi toutes les intersections ont un comportement conjonctif.

Il existe aussi des opérateurs dont le comportement est dynamique suivant ses arguments. Parfois le résultat est un compromis, parfois c'est une union...

Trois exemples suivent:

$$\Gamma_{\gamma}(x,y) = [\cap(x,y)]^{\gamma} [\cup(x,y)]^{1-\gamma} \quad (1.16)$$

$$A_{\gamma}(x,y) = \gamma[\cap(x,y)] + (1-\gamma)[\cup(x,y)] \quad (1.17)$$

Soit une fonction $g: [0,1]^2 \rightarrow [0,1]$ telle que $g(0,0)=0$, g croissante et continue par rapport à chacun de ses arguments, alors il est possible de définir une somme symétrique [Silvert 1979]:

$$\sigma(x,y) = \frac{g(x,y)}{g(x,y) + g(1-x,1-y)} \quad (1.18)$$

Par exemple pour $g(x,y)=\min(x,y)$, la somme suivante est obtenue:

$$\sigma_{\min}(x,y) = \frac{\min(x,y)}{1-|x-y|} \quad (1.19)$$

Pour $\gamma \in]0,1[$ les opérateurs Γ et A ont toujours un comportement de compromis. Le comportement d'une somme symétrique dépend de ses arguments.

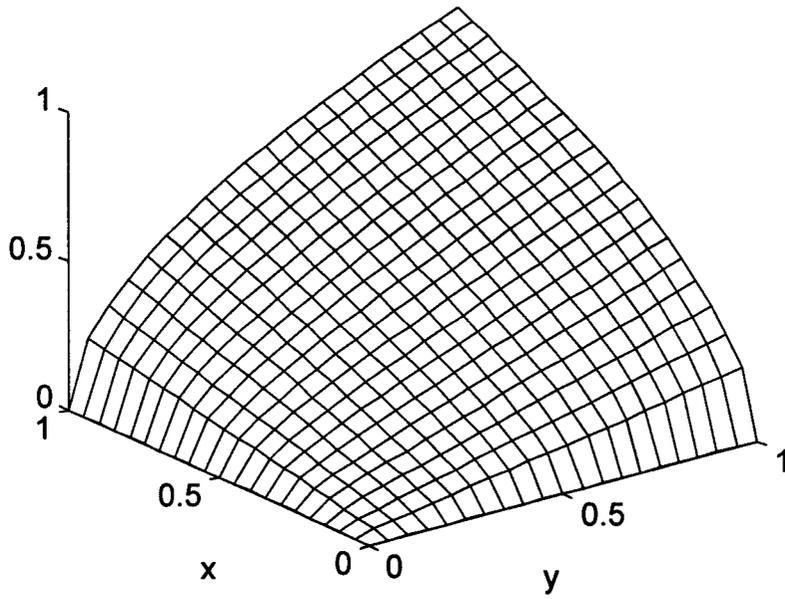


Fig. 1.16. $\Gamma_{0.5}(x,y)$ pour la famille probabiliste.

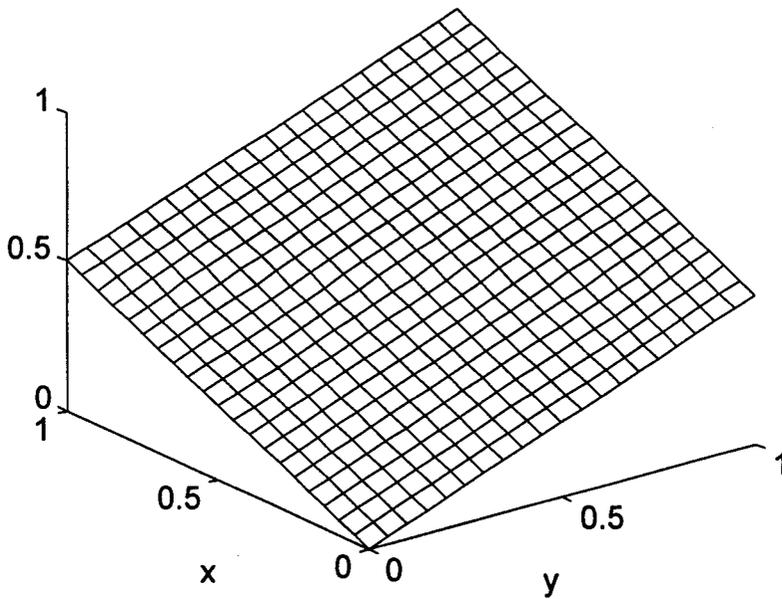


Fig. 1.17. $A_{0.5}(x,y)$ pour la famille probabiliste.

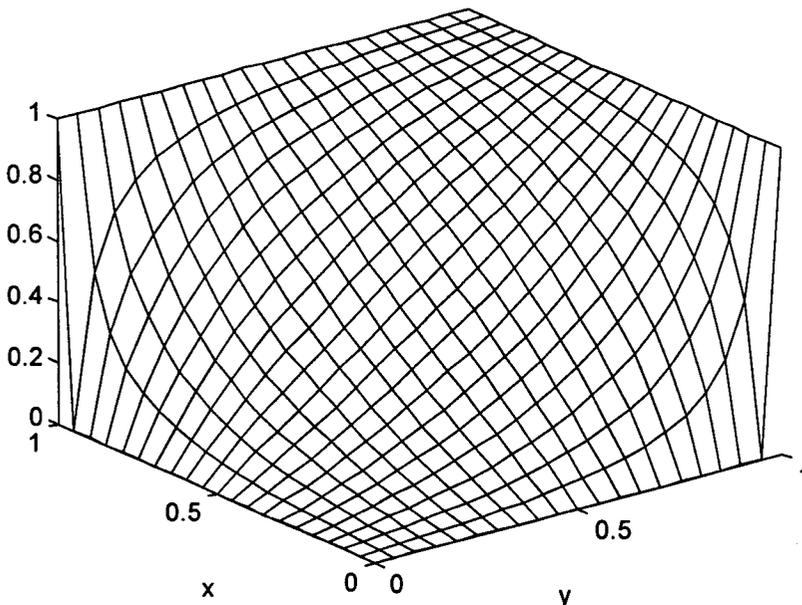


Fig. 1.18. $\sigma_{\text{prod}}(x, y) = \frac{xy}{2xy + 1 - x - y}$.

1.4 Des ensembles flous à la logique floue

Les ensembles flous viennent d'être définis, ainsi que les opérations élémentaires d'union, d'intersection, et de complément.

Le principe de la logique booléenne est de lier des notions par des relations de type implication, ou plus précisément de lier des faits à des conclusions. Par exemple pour régler la température de l'eau du bain, si la température est froide (c'est une mesure qui donne un fait), alors il faut ajouter de l'eau chaude (c'est la partie conclusion).

En logique booléenne, toute implication peut être représentée comme une relation, ou encore un ensemble, défini sur le produit cartésien des espaces de référence servant à définir les prémisses et les conclusions de l'implication. Par exemple dans le réglage de la température du bain, l'espace de référence de la prémisse est la température de l'eau, l'espace de référence de la conclusion est également la température de l'eau (le même que celui de la prémisse, mais de manière générale il peut être différent). Et l'implication représente bien une relation entre les deux espaces de référence: Si la température est froide, alors il faut ajouter de l'eau chaude, et le graphe de la relation utilise les deux espaces de référence.

La logique floue réalise exactement la même fonction: lier des prémisses à des conclusions. Cependant la logique floue a l'avantage de pouvoir aborder des notions imprécises, floues, vagues, celles-ci étant traitées par les ensembles flous. Ainsi l'exemple donné du réglage de la température de l'eau relève plus de la logique floue que de la logique classique, à moins d'assigner des valeurs très précises aux concepts de froid et chaud (par exemple 20°C pour le froid, et 45°C pour le chaud), le capteur étant binaire et ne renvoyant de même que des températures très précises (soit 20°C, soit 45°C).

Dans la réalité les températures ne sont pas précisément définies et les capteurs renvoient des mesures imprécises. Les chapitres qui suivent montrent comment la logique floue traitera le problème.

D'abord, comme toute implication, booléenne ou floue, est une relation définie sur des produits cartésiens d'ensembles, quelques notions sur les ensembles flous seront données. Puis l'implication floue sera expliquée, et enfin sera généralisé le Modus Ponens, paradigme du raisonnement et de la logique.

1.4.1 Projection, Produit Cartésien et Extension

Etant donné un ensemble flou A défini sur un référentiel $X \times Y$, il est possible de projeter A sur X ou sur Y. Par exemple la projection de A, noté $A(X \times Y)$ sur X, notée $A(X)$, est obtenue par:

$$\mu_{A(X)}(x) \hat{=} \max_Y (\mu_{A(X \times Y)}(x, y)) \quad (1.20)$$

Etant donné un ensemble flou A défini sur X, noté $A(X)$, et un ensemble B défini sur Y, noté $B(Y)$, un ensemble flou $A \times B$ défini sur $X \times Y$ peut être obtenu par:

$$\mu_{A \times B(X \times Y)}(x, y) \hat{=} \min(\mu_{A(X)}(x), \mu_{B(Y)}(y)) \quad (1.21)$$

Etant donné un ensemble flou A défini sur X, noté $A(X)$, il est possible d'en déduire un ensemble flou A étendu à $X \times Y$, noté $A(X \times Y)$, par la relation:

$$\forall (x, y) \in X \times Y, \mu_{A(X \times Y)}(x, y) \hat{=} \mu_{A(X)}(x) \quad (1.22)$$

Ces trois opérations sont en pratique importantes. Elles sont utilisées notamment dans le principe du raisonnement flou, et aussi lorsque sont affrontés des problèmes où de multiples attributs sont présents.

Par exemple si une population est étudiée, et qu'un ensemble flou résume toute l'information disponible en en qui concerne la couleur des yeux et la couleur des cheveux, il sera possible d'étudier séparément la couleur des yeux de la population en projetant l'ensemble flou représentant la connaissance sur l'attribut (ou l'espace de référence) 'couleur des yeux'.

1.4.2 Implication floue

Une implication est un lien logique qui lie deux variables: l'observation d'une prémisse entraîne obligatoirement la conclusion de l'implication. Le tableau 1.2 montre la table de vérité de l'implication booléenne.

	y	0	1
x		0	1
0		1	1
1		0	1

Tableau 1.2. Table de vérité de $x \Rightarrow y$.

En fait l'idée principale est qu'une implication (à distinguer de l'équivalence) est toujours vraie sauf lorsque la prémisse est vraie et que la conclusion est fausse (ce qui évidemment est contraire à l'implication).

Avec les ensembles flous, les valeurs des ensembles ne sont plus binaires. Apparaît alors un problème lorsque la généralisation de l'implication est abordée. Ainsi dans quelle mesure peut il être affirmé que l'implication est fausse lorsque la prémisse est à moitié vraie et la conclusion à moitié fausse? De même l'implication doit-elle être complètement fausse dès que

la valeur d'appartenance de x est supérieure à celle de y, ou seulement être presque fausse, avec un ensemble flou correspondant proche de 0? Aucune réponse ne peut venir de la logique classique, et plusieurs interprétations de l'implication floue ont été proposées. Trois grandes familles peuvent être définies [Lee 1990, part 2; Bezdek, 1986; Bouchon-Meunier, 1993] :

1) Basée sur la généralisation de l'expression classique de l'implication

$$(A \Rightarrow B) \Leftrightarrow (\text{Non } A \text{ Ou } B)$$

Etant donné une implication et une union, l'expression est

$$\mu_{\Rightarrow}(x,y) = \cup(n(\mu_A(x), \mu_B(y)))$$

2) L'expression générale de l'implication est:

$$\mu_{\Rightarrow}(x,y) = n(\cap(\mu_A(x), n(\cap(\mu_A(x), \mu_B(y))))))$$

Elle représente la généralisation de la relation $(A \Rightarrow B) \Leftrightarrow [\text{Non } A \text{ Ou } (A \text{ Et } B)]$

(remarque: cette expression est différente de celle de la première classe, puisqu'en logique floue, "A Ou Non A" est différent de la tautologie). L'implication 2 en est un exemple.

3) L'expression générale de l'implication est:

$$\mu_{\Rightarrow}(x,y) = \cap^{-1}(\mu_A(x), \mu_B(y))$$

avec $\cap^{-1}(x,y) = \max(z: \cap(x,z) \leq y)$

Le principe de cette implication est le suivant: si $\mu_A(x) \leq \mu_B(x)$, alors la valeur de l'implication est égale à 1. Par contre si $\mu_A(x) \geq \mu_B(x)$, la valeur de l'implication peut être beaucoup plus faible, au pire être égale à $\mu_B(x)$, selon l'intersection choisie. Si l'intersection choisie est le min, alors quelque que soit la valeur de $\mu_A(x) \geq \mu_B(x)$, le résultat est $\mu_B(x)$, et si $\mu_B(x) = 0$, l'implication est complètement fausse.

Les implications 5 et 6 en sont des exemples.

Des implications floues particulières sont maintenant données:

$$1) \mu_{A \Rightarrow B}(x,y) = 1 - \mu_A(x) + \mu_A(x)\mu_B(y)$$

$$2) \mu_{A \Rightarrow B}(x,y) = \max(1 - \mu_A(x), \min(\mu_A(x), \mu_B(y)))$$

$$3) \mu_{A \Rightarrow B}(x,y) = \min(1 - \mu_A(x) + \mu_B(y), 1)$$

$$4) \mu_{A \Rightarrow B}(x,y) = \begin{cases} 1 & \text{si } \mu_A(x) \leq \mu_B(y) \\ 0 & \text{sin on} \end{cases}$$

$$5) \mu_{A \Rightarrow B}(x,y) = \begin{cases} 1 & \text{si } \mu_A(x) \leq \mu_B(y) \\ \mu_B(y) & \text{sin on} \end{cases}$$

$$6) \mu_{A \Rightarrow B}(x,y) = \begin{cases} \min(1, \mu_B(x) / \mu_A(y)) & \text{si } \mu_A(x) \neq 0 \\ 1 & \text{sin on} \end{cases}$$

$$7) \mu_{A \Rightarrow B}(x,y) = \mu_A(x) \cap \mu_B(y) \text{ (une intersection utilisée en commande floue)}$$

Ces sept implications lient deux ensembles flous A (défini sur X) et B (défini sur Y), en traduisant l'implication $A \Rightarrow B$.

Elles sont toutes identiques à la logique classique dans le cas où A et B sont des ensembles classiques, à part la dernière qui n'est qu'une intersection. Elle n'est utilisée qu'en automatique pour sa simplicité, et surtout parce que les prémisses sont considérées comme les entrées de

systèmes à commander, et que les conclusions, étant considérées comme des sorties, ne sont jamais indépendantes des entrées.

La figure 1.19 montre un exemple d'implication dans le cas où $\mu_A(x)=x$ et $\mu_B(y)=y$.

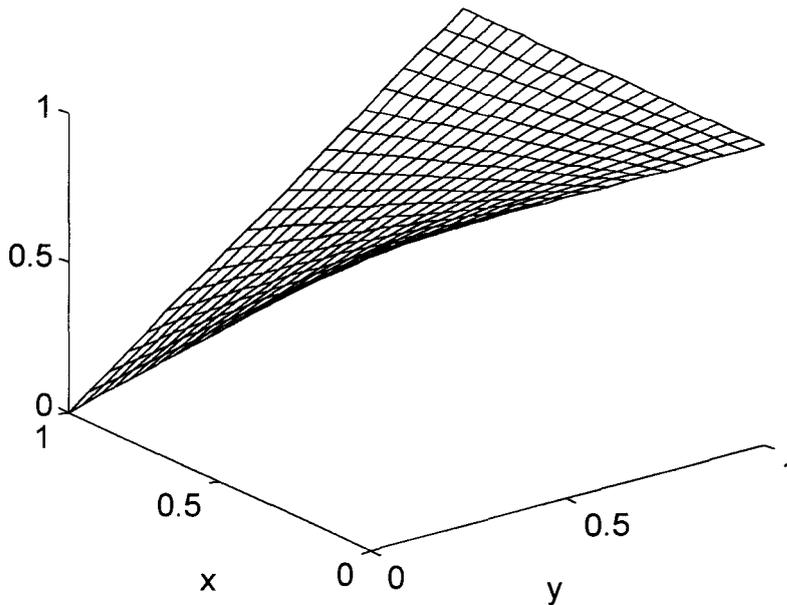


Fig. 1.19. $x \Rightarrow y$ représenté par $1-x+xy$.

Les quatres sommets de la figure ($x=0, y=0, (x \Rightarrow y)=1$; $x=1, y=1, (x \Rightarrow y)=1$; $x=0, y=1, (x \Rightarrow y)=1$ et enfin $x=1, y=0, (x \Rightarrow y)=0$) sont identiques au cas classique. Cependant la figure apparaît déformée dans l'espace, et n'appartient pas au plan des sommets.

Les six premières implications (en fait les seules vraies généralisations), peuvent être classées en trois familles

1.4.3 La logique floue: Le Modus Ponens généralisé

1.4.3.1 Le principe

Les raisonnements en logique classique sont basés sur quelques modèles de raisonnement remarquables. Le plus classique d'entre eux est le Modus Ponens. Il est encore appelé syllogisme et s'énonce ainsi:

- Connaissance a priori: $A \Rightarrow B$ (prémisse majeure)
- Observation (fait) de A (prémisse mineure)
- Donc B est induit (conclusion)

L'archétype du raisonnement est le célèbre 'Tous les hommes sont mortels. Socrate est un homme. Donc Socrate est mortel'.

Ce principe de raisonnement est mis en défaut lorsque des faits A' légèrement différents des prémisses des implications sont observés. Dans ce cas, la logique classique ne peut rien déduire.

Ainsi si la prémisse majeure était que tous les singes étaient mortels, rien n'aurait pu être dit sur la capacité de Socrate à être mortel.

C'est pourquoi les ensembles flous, tout en nuances, permettent d'enrichir ce principe de raisonnement. Si des faits sont légèrement différents des prémisses d'une implication, une certaine information pourra en être déduite. Ceci est le Modus Ponens généralisé.

- Connaissance a priori: $A \Rightarrow B$
- Observation (fait) de A'
- Donc B' est induit

La procédure est la suivante:

La connaissance a priori est représentée par l'implication entre A et B , c'est à dire une relation R , définie sur le produit cartésien des espaces de références X et Y de A et B ($A \subset X, B \subset Y$): $R \subset X \times Y$.

Le seul fait à disposition est un ensemble flou A' défini sur X .

Il va falloir le comparer à la connaissance a priori R .

La première étape consiste à étendre l'ensemble flou A' sur $X \times Y$ de manière à pouvoir le comparer à R . Ceci est réalisé par une extension cylindrique (1.20).

Cette opération est transparente, puisque la fonction d'appartenance n'est pas modifiée numériquement.

La seconde étape consiste à comparer l'ensemble flou A' étendu à R . Ceci est réalisé par une intersection. Le résultat est un ensemble flou B' étendu à l'espace $X \times Y$.

Pour avoir une conclusion définie sur Y , il suffit alors de le projeter sur l'espace Y , par la relation 1.22.

Ces différentes étapes peuvent être résumées par la relation:

$$\mu_{B'}(y) = \sup_x (\mu_{A'}(x), \mu_{A \Rightarrow B}(x, y)) \quad (1.23)$$

En résumé le Modus Ponens généralisé fait intervenir trois fonctions: d'abord pour construire la relation R , il faut choisir un opérateur d'implication. Ensuite, il faut choisir un opérateur d'intersection pour comparer les faits aux connaissances a priori. Enfin il faut projeter le tout sur l'espace des conclusions. La quatrième opération, l'extension, ne fait intervenir aucune fonction particulière, et n'apparaît pas dans la relation 1.23.

Un modèle de raisonnement pourrait être le suivant: "Tous les singes sont mortels. Socrate est un homme. Donc Socrate est (presque certainement) mortel".

Cet exemple soulève plusieurs problèmes: D'abord il faudrait trouver un espace de référence sur lequel seraient définis les entités singe et homme. Il pourrait s'agir d'un produit cartésien de plusieurs attributs, par exemple de morphologie, de capacité à parler...

Ensuite il apparaît que la conclusion a été modifiée. Ainsi est apparu un "presque certainement". En fait en logique floue, suivant le choix particulier des opérateurs d'intersection et d'implication, les résultats apparaissent plus ou moins modifiés par rapport à la conclusion normale. Les paragraphes qui suivent donnent quelques règles d'analyse.

1.4.3.2 Compatibilité entre les opérateurs

Si les faits observés A' sont identiques aux véritables prémisses, il serait souhaitable que les conclusions B' soient aussi identiques aux véritables conclusions B .

Donc pour le choix des opérateurs d'implication et d'intersection, la contrainte suivante est ajoutée:

$$(A'=A) \Rightarrow (B'=B)$$

En fait il y a une certaine restriction entre les opérateurs, comme le tableau 1.3 le montre:

Choix de l'intersection \cap du Modus Ponens	Implication \Rightarrow Imposée
$\cap = \max(x+y-1, 0)$	1, 2, 3, 4, 5, 6, 7
$\cap = \min(x, y)$	4, 5, 7
$\cap = xy$	4, 5, 6, 7
toute \cap	\Rightarrow construite par la troisième classe (cf 1.4.2)

Tableau 1.3. Contrainte entre les opérateurs pour le Modus Ponens généralisé.

1.4.3.3 Etude de l'évolution de la conclusion

La conclusion du petit exemple du paragraphe précédent était 'Socrate est (presque certainement) mortel'.

L'étude des modifications apportées aux conclusions lorsque les faits disponibles sont différents des prémisses est très intéressante. Son étude complète nécessiterait un grand nombre de pages, mais ce paragraphe donne quelques résultats fondamentaux.

La première analyse possible est celle du cas où l'observation est plus précise que la prémisse. En fait l'observation est un sous ensemble de la prémisse, et elle est complètement incluse dedans: $A' \subset A$.

Logiquement la conclusion devrait être identique ($B' = B$), car avec une seule implication, il n'y a absolument aucun renseignement sur les évolutions corrélées des prémisses et des conclusions. Le fait qu'un homme soit blond ou jeune n'améliore pas sa capacité à être mortel (peut être son espérance de vie, mais il faut alors d'autres implications modélisant ces rapports). En fait l'information supplémentaire devrait être logiquement abandonnée, et être perdue.

Avec le Modus Ponens généralisé, cette règle est respectée, sauf lorsque l'implication 6 est choisie. Avec l'implication 4, la conclusion B' est aussi plus précise que la véritable conclusion B . Ceci est un comportement aberrant et discutable.

La seconde analyse possible concerne le cas où les faits A' sont moins précis que les prémisses A des règles. Les prémisses sont incluses dans les faits: $A \subset A'$. Ce cas est complexe, et plusieurs sous problèmes apparaissent.

Dans le cas où A et A' ont le même cœur, alors les conclusions B et B' ont aussi le même cœur.

Dans le cas où A et A' ont simplement le même support, des différences apparaissent suivant le choix d'une implication. En règle générale la conclusion modifiée contient la véritable conclusion $B' \supset B$.

- Pour les implications 4, 5, 6, 7, les conclusions B et B' ont le même support
- Pour les implications 1, 2, 3, B' est identique à B à part un certain niveau d'incertitude ε (voir la figure 1.20). ε dépend des ensembles flous A et A'

Dans le cas où A est simplement inclus dans A' , sans aucune information sur le cœur ou le support, en règle générale B' contient B .

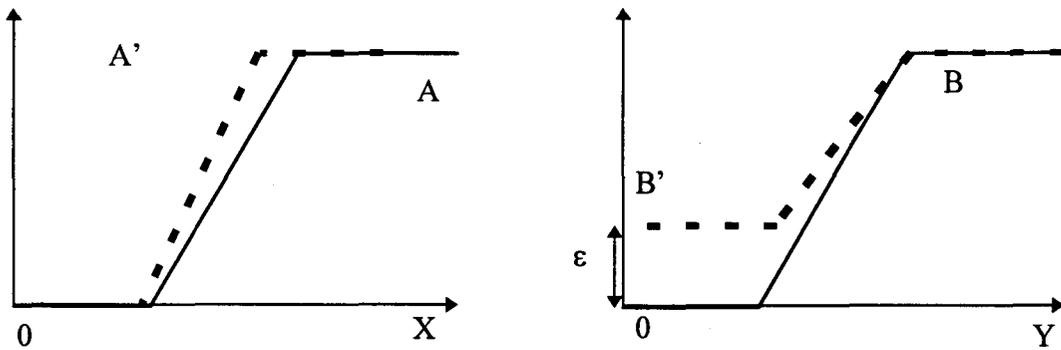


Fig. 1.20. Etude de la conclusion du Modus Ponens généralisé lorsque la prémisse de la règle A est contenue dans un fait A'. A et A' ont le même support. L'implication retenue est la (1).

1.4.3.4 Exemple de calcul d'une implication

La simplicité de cet exemple permettra de montrer les mécanismes mis en jeu par la logique floue.

La figure 1.21 va donner les deux labels caractérisant les référentiels.

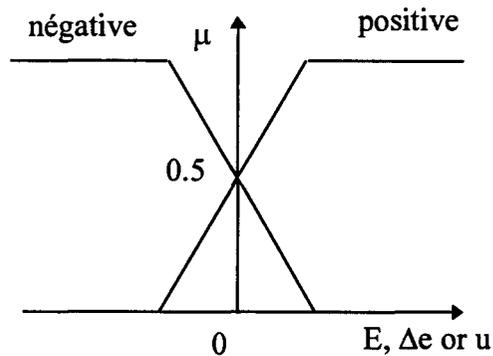


Fig. 1.21. Les labels caractérisant le référentiel.

La règle est:

si l'erreur E est positive et Δe est positive alors la commande u est négative

Maintenant, que se passe-t-il si E est E_1 et Δe est Δe_1 ?

L'inférence va être basée sur une logique simple:

\Rightarrow est une t-norme (min)

\cap est le min (Inférence basée sur la famille min-max)

La figure 1.22 fournit les nouvelles valeurs E_1 et Δe_1 :

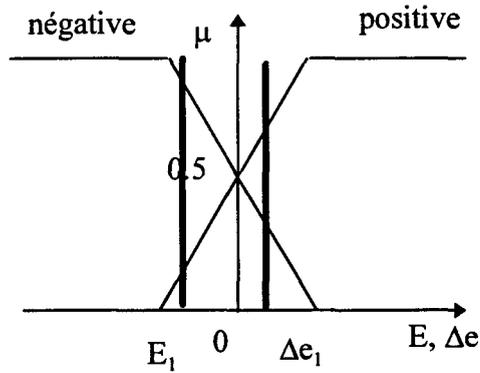


Fig. 1.22. Les entrées.

Il faut maintenant faire la comparaison des entrées aux anciens ensembles flous. La figure 1.23 donne les intersections, qui sont simples à calculer puisque les nouvelles mesures sont des singletons flous.

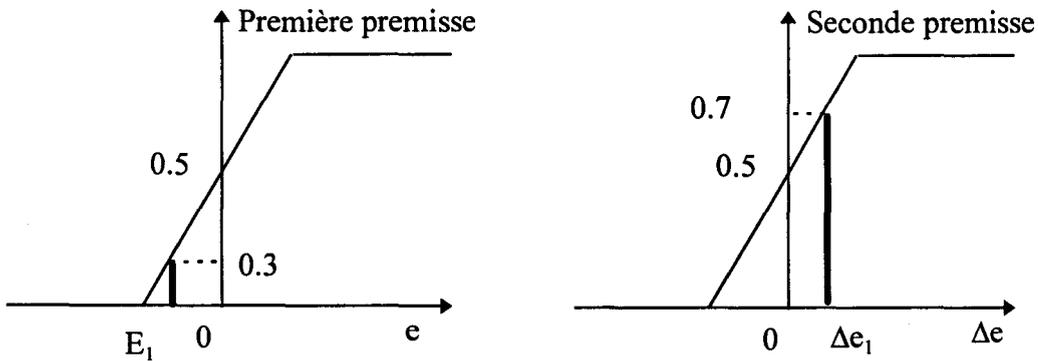


Fig. 1.23. Les entrées comparées aux prémisses de la règle.

La valeur minimale entre ces deux intersections est 0.3, et donc il est possible de calculer l'ensemble flou de la conclusion, donné à la figure 1.24:

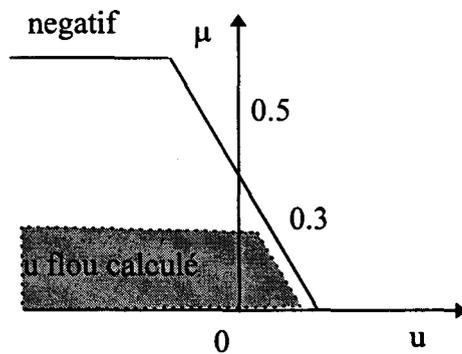


Fig. 1.24. Conclusion obtenue à partir de la règle floue et des entrées.

Dans cet exemple, la conclusion est un sous ensemble de la véritable conclusion, mélangée avec d'autres règles, elle aurait donc moins de poids que si les entrées respectaient parfaitement les prémisses.

1.5 Conclusion

Ce chapitre a introduit les ensembles flous par rapport aux ensembles classiques. Les opérations élémentaires d'union, d'intersection et de négation ont également été définies. Dans la continuité de la théorie classique, l'implication logique fut généralisée, amenant naturellement au raisonnement flou.

Il apparaît que les ensembles flous apportent à la théorie classique la capacité de représenter mathématiquement les concepts vagues et mal définis. Ils donnent une souplesse certaine au raisonnement logique en général, et apparaissent finalement comme un outil judicieux pour traiter des données imprécises et mimer le raisonnement humain face aux systèmes complexes.

Beaucoup de champs d'application possibles de la théorie des ensembles flous n'ont pas été abordés. Ainsi il est possible de définir une mesure floue du degré d'inclusion d'un ensemble dans un autre. Le principe d'extension permet de calculer l'image d'un ensemble flou par une fonction quelconque. Il est de même possible de définir une arithmétique floue basée sur des nombres flous, ensembles flous auxquels a été ajoutée, entre autres, une propriété de convexité.

Les deux chapitres qui suivent reviennent sur deux aspects importants qui sont issus de la logique floue. Le premier concerne la théorie des possibilités. Il s'agit d'une interprétation des degrés d'appartenance en termes d'imprécision et d'incertitude, l'ensemble de la théorie des possibilités donnant tous les outils nécessaires à la représentation et au traitement logique des données incertaines.

Le second chapitre aborde de manière détaillée le contrôleur flou. A l'image des systèmes experts flous, il utilise le raisonnement flou, avec en outre la faculté de considérer un grand nombre d'implications en parallèle, mais il est dédié au champ de la commande.

Ces outils flous, la théorie des possibilités, et la généralisation du contrôleur flou, amèneront dans la suite de cette thèse au contrôleur multi-modèle.

CHAPITRE DEUXIEME

LA THEORIE DES POSSIBILITES

2.1 Introduction: Les données incertaines

Ce chapitre est dédié aux traitements logiques qui peuvent être effectués à partir de données incertaines. Avant de les manipuler, il faudra d'abord modéliser ces informations imparfaites, par une théorie appropriée, il s'agira ensuite d'éliminer ces imperfections pour obtenir des données propres à être utilisées par d'autres outils mathématiques, ou encore pour obtenir des données certaines qui pourront être utilisées en toute quiétude, sans avoir à se demander si elles sont fausses ou non.

Il ne s'agit en aucun cas de filtrage, par exemple, mais simplement d'une méthode visant à éliminer les imperfections par la prise en compte simultanée de plusieurs sources d'information, de manière à corréler (sans aucun sens statistique) les données pour en ressortir la "substantifique moelle".

Avant d'aborder les outils gérant les données imparfaites, il faut donc s'entendre sur le sens qui est donné aux mots incertitude, imperfection,...

Lorsqu'une information est fournie par un capteur, ou toute autre source d'information comme un témoignage humain par exemple, plusieurs problèmes relatifs à la qualité des données fournies apparaissent. Il est possible de les classer en trois grandes catégories.

Le premier problème est celui de la précision des informations fournies. En effet une mesure donnée par un capteur peut ne pas être stable dans le temps, et osciller légèrement. Les mesures peuvent être perturbées par toutes sortes de bruits. Les capteurs peuvent ne pas être aussi sensibles qu'ils le devraient. Par exemple en robotique lorsqu'il s'agit de diriger un robot, plusieurs types de capteurs peuvent être utilisés. Des mesures de distance peuvent être données par des télémètres lasers, mais les appareillages coûtent alors très cher, et parfois il faut se rabattre sur des capteurs moins coûteux, par exemple des capteurs à ultrasons. Ceux-ci fournissent une mesure de distance des obstacles, mais avec un cône de détection de 30°...

Dans cette classe de problèmes liés à la précision des informations, il est aussi possible d'associer les informations fournies par des opérateurs humains, telle l'eau qui est chaude, concept ô combien subjectif et donc imprécis.

La seconde classe d'incertitudes liée aux informations est celle de la crédibilité ou de la confiance que l'on accorde aux mesures. Par exemple les matériels ont une certaine durée de vie, et si un capteur a un temps moyen avant défaillance de 5 ans, et qu'il est utilisé depuis 10 ans, il devient alors moins crédible, car il pourra alors être éventuellement défaillant. De même un capteur peut éventuellement être contrôlé à certains intervalles de temps, et l'utilisateur aura plus confiance dans la fiabilité du capteur juste après un contrôle, que juste avant, surtout si les périodes entre les contrôles sont longues.

Il apparaît encore dans cette classe le facteur humain. C'est peut être l'emblème le plus révélateur des difficultés soulevées par des informations non fiables.

Ainsi il y a quelques siècles, les rapports des magistrats concernant des affaires judiciaires disaient ainsi que le témoignage de la voisine d'une amie d'un témoin mort par la suite était considéré comme 1/12 de preuve...

Aujourd'hui il existe une théorie des témoignages, et il a été montré que des témoins de toute bonne foi, et sincères, pouvaient se tromper complètement. D'ailleurs les problèmes judiciaires sont traités préférentiellement lorsqu'il existe des preuves matérielles. Mais parfois le

facteur humain est indispensable, et certains humains sont considérés comme plus fiables que d'autres.

La troisième grande classe de problèmes liée aux informations fournies, est moins liée à leur nature même, qu'à leur quantité. Il s'agit tout simplement de la complétude des informations disponibles.

Ainsi il serait difficile de respecter les limitations de vitesse si les voitures n'étaient pas équipées de compteur de vitesse. Cet exemple trivial ne doit pas masquer le fait qu'il peut être nécessaire dans des situations complexes de devoir prendre une décision alors que certains paramètres cruciaux ne sont pas connus.

Il arrive alors que des références plus "universelles" soient utilisées. Universelles dans le sens que ce sont des attitudes ou des choix utilisés indépendamment des problèmes. Ainsi en analyse de la décision, une hypothèse optimiste ou pessimiste quant à la nature des probabilités de certains événements est utilisée en dernier ressort. Optimiste si la nature est perçue comme agissant dans le sens de l'utilisateur, pessimiste dans le cas contraire. Suivant le choix retenu, sont utilisés des algorithmes de type Maximax ou Maximin...

Ces trois grandes catégories d'imperfections recouvrent la majeure partie des incertitudes liées aux informations. Suivant les sources de données, ces quantités évoluent souvent de manière antagoniste.

L'exemple qui suit montre ces évolutions contraires pour la précision et la confiance qui sont accordées aux informations.

Il s'agit d'estimer la longueur d'une voiture. Quatre sources d'informations sont possibles: un fou, les capteurs visuels de celui qui a posé la question (autrement dit, ses yeux!), sa culture générale, et enfin le manuel technique de la voiture.

précision certitude	petite précision	grande précision
faible crédibilité	entre 1 et 10 km (fournie par le fou) <u>Donnée inutile</u>	4,44 mètres (fourni par les yeux)
grande crédibilité	entre 0 et 10 mètres (culture générale)	<u>Donnée idéale</u> 5,79 mètres (fournie par le constructeur)

Fig. 2.1. Précision contre certitude.

Les données les plus intéressantes sont celles qui sont à la fois précises et certaines (c'est à dire avec un haut degré de fiabilité). Les données peu précises et peu certaines sont le moins utile, et entre les deux il existe toute une série de cas intermédiaires.

Il existe aussi des capteurs intelligents, comme les humains, qui peuvent adapter eux mêmes la qualité de l'information fournie en fonction de la précision requise: "Quel âge a cet homme?" demande A. "Oh, la trentaine", répond B. "En êtes vous sur?", demande A. "Absolument répond" B. "Mais ne pourriez vous être plus précis ?", demande A. "Il doit avoir 32 ans, mais je n'en suis pas sûr" répond finalement B.

Après avoir présenté les données imparfaites, il s'agit maintenant de présenter les outils les maniant.

La première théorie est celle des probabilités, qui est unanimement reconnue dans le monde scientifique. Afin de déterminer la capacité des événements à se réaliser, sont attachés à ces derniers des degrés de probabilité. Suivant la nature des données utilisées pour calculer ces degrés, peuvent être définies trois familles de probabilités. La première concerne le cas exact du dénombrement. Cependant en général les problèmes sont trop complexes pour la combinatoire, et la deuxième famille comprend les probabilités statistiques liées à des répétitions d'expérience aléatoires, normalement en nombre infini. Dans le langage courant, le terme probable est utilisé pour établir le fait qu'un événement a de très grandes chances de se réaliser, mais en dehors de toute notion statistique. C'est pourquoi fut introduite la dernière famille des "probabilités subjectives", pour traiter des degrés fournis par une expérience humaine, avec un formalisme utilisant des paris.

Cependant même ces trois types de probabilités peuvent se révéler inutiles dans certains cas. Soit il n'est pas possible de définir les degrés de probabilités parce que le nombre d'expérience est insuffisant (pour les probabilités statistiques). Soit la contrainte d'additivité est trop contraignante pour les probabilités subjectives. Soit, l'impossibilité de pouvoir représenter l'ignorance, totale ou partielle, due à une *obligation* de quantifier exactement les degrés de probabilités pour chaque événement, empêche l'utilisation des probabilités, qui, sinon, transformeraient les problèmes en exigeant l'introduction artificielle de données.

En 1976 apparut la théorie de Dempster Shafer. Plusieurs interprétations peuvent en être tirées, mais une simple se réfère aux probabilités, qui sont alors généralisées [Smets, 1993, pour les différentes familles]: il suffit alors d'imaginer qu'en cas de données incertaines ou incomplètes, les degrés de probabilités sont remplacés par des intervalles de probabilités. Au lieu d'avoir une mesure de probabilité, deux nouvelles fonctions apparaissent: une borne inférieure, appelée mesure de crédibilité ou de croyance, et une borne supérieure, appelée mesure de plausibilité. Lorsque les informations sont suffisantes, les deux fonctions sont égales, et les fonctions de croyance dégénèrent en mesures de probabilités.

La théorie de Dempster Shafer, encore appelée théorie des preuves ou de l'évidence, a des liens très forts avec les probabilités, cependant ses axiomes principaux sont plus généraux. Ainsi la propriété d'additivité des probabilités ($P(A \cup B) = P(A) + P(B)$ si l'intersection, est vide) est perdue pour les deux mesures, de croyance, et de plausibilité.

Bien qu'étant complexes à manier, et nécessitant beaucoup de calculs, les fonctions de croyance constituent un cadre très général pour aborder les données incertaines. Aujourd'hui il existe plusieurs écoles d'interprétations de la théorie de Dempster Shafer, et l'une d'entre elles utilise même les deux mesures de croyance et de plausibilité comme des indicateurs de sentiments humains, en dehors de toute référence probabiliste.

En 1978 [Zadeh, 1978] fut créée la troisième grande théorie de l'incertitude, sous le nom de théorie des possibilités. Dérivée des ensembles flous, elle est beaucoup moins complexe à manier que les fonctions de croyance. Pourtant elle permet de représenter des notions comme

l'imprécision et l'incertitude ou la fiabilité des données. Contrairement à la théorie de Dempster Shafer, elle a peu de liens avec les données aléatoires, et ses axiomes de base sont complètement différents de ceux des fonctions de croyance ou des probabilités.

Dans ces trois grandes théories, l'utilisateur qui manipule des informations incertaines cherche d'abord à les rendre un peu plus utilisables au point de vue de la fiabilité et de la précision. Si certaines données sont incomplètes, l'utilisateur cherche aussi à les compléter. Pour cela, le principal moyen est la mise en parallèle ou la fusion de différentes sources d'information. C'est un principe de bon sens que de chercher à utiliser une autre information si une source de données est défaillante.

Par conséquent, dans ces trois théories, des outils permettant la fusion d'informations ont été développés. Le multi-modèle présenté à la fin de cette thèse utilise ces outils. Considérant des structures élémentaires fournissant des informations, encore appelés modèles, l'objectif du multi-modèle est de les fusionner pour obtenir l'information la plus sûre possible.

Aussi la présentation de ces outils de fusion prendra une place importante dans ce chapitre. Bien qu'elles aient été étudiées en détail, les fonctions de croyance n'ont pas encore trouvées d'application avec le contrôleur multi-modèle envisagé. C'est pourquoi elles ne seront pas présentées dans ce chapitre.

Dans une première partie la définition des mesures de possibilités sera donnée. Celle-ci sera introduite par les liens des possibilités avec respectivement les ensembles flous et les probabilités. Quelques résultats théoriques et les axiomes fondamentaux des possibilités seront alors présentés.

Ensuite un paragraphe important exposera les principales règles de fusion utilisées dans cette théorie. Leurs hypothèses d'application, leur avantages et leurs inconvénients seront alors montrés.

Finalement un nouvel opérateur de fusion, permettant la fusion en parallèle de multiples sources d'information, chacune avec un degré de fiabilité propre, sera mis en avant. Quelques analyses et résultats issus de ce nouvel opérateur seront finalement proposés.

2.2 Définitions et propriétés des possibilités

Avant de donner quelques relations mathématiques, il faut d'abord s'entendre sur le sens donné à une mesure de possibilité. Un degré de possibilité mesure la capacité d'un événement à se produire ou à être. C'est une mesure d'une capacité physique à exister, indépendamment d'autres notions statistiques par exemple. Les possibilités généralisent de même le calcul d'erreur utilisé en physique, en nuancant les termes de possible et nécessaire [Dubois et Prade, 1988]. Les deux exemples qui suivent montrent les liens de ces degrés de possibilités avec les ensembles flous et les probabilités.

2.2.1 Liens entre les possibilités et les ensembles flous

Pour résumer en quelques mots le premier chapitre, les ensembles flous sont relatifs à l'appartenance d'éléments bien précis à des ensembles mal définis.

La figure 2.2 rappelle comment des concepts humains peuvent être représentés par des ensembles flous.

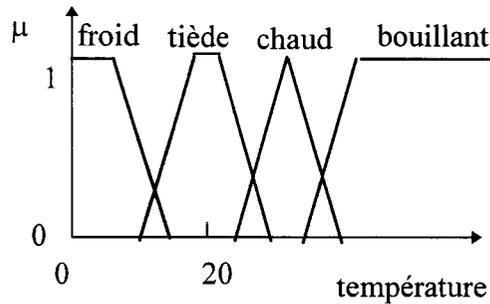


Fig. 2.2. Ensembles imprécis caractérisant la température de l'eau.

En fait la fonction d'appartenance μ permet de caractériser dans quelle mesure un élément précis appartient à un ensemble mal défini.

Autrement dit elle permet de savoir si un élément très précis, par exemple 21,6° Celsius, fait partie d'un ensemble mal défini, par exemple 'tiède', ou s'il n'en fait pas partie, et dans quelle proportion.

Un degré de possibilité procède d'une approche exactement contraire [Dubois et Prade, 1982; Zadeh, 1978].

Etant donné un objet mal défini, imprécis, incertain, la question est maintenant d'estimer dans quelle mesure il peut être égal à tel ou tel élément très précis.

Par exemple, étant donné l'objet mal défini 'tiède', est-il possible qu'il représente une température de 29° Celsius, et est-il possible d'évaluer la véracité de cette égalité?

Si une distribution de possibilité $\pi_A: X \rightarrow [0,1]$ (X étant l'espace de référence) est introduite pour représenter des degrés de possibilité d'éléments de X par rapport à un ensemble A , il est logique de penser que cette distribution de possibilité ait des rapports avec une fonction d'appartenance $\mu_A: X \rightarrow [0,1]$ caractérisant l'ensemble A .

En fait, Zadeh a posé la relation suivante, qui énonce que les rapports sont de type égalité:

$$\forall x \in X, \pi_A(x) = \mu_A(x) \quad (2.1)$$

Cette relation est appelée principe possibiliste.

Cette relation pourrait être réécrite de la manière suivante:

$$\forall x \in X, \pi(x|A) = \mu(A|x) \quad (2.2)$$

Ainsi il pourrait être plus facile de comprendre les objets manipulés par les deux théories [Smets, 1993].

2.2.2 Liens avec les probabilités

Comme il a été dit au début de ce paragraphe, les possibilités mesurent les capacités physiques d'événements à se réaliser. Elles ne sont pas liées à des connaissances statistiques. Cependant l'exemple qui suit montre les liens entre une distribution de probabilité p et une distribution de possibilité π .

Quel est le nombre de verres de cognac que vous pouvez boire le soir en rentrant du travail, tout en restant lucide (c'est à dire en étant encore capable de mettre votre index sur le bout de votre nez, avec les yeux fermés)?

Pour répondre à cette question, deux distributions, une de probabilité et une de possibilité, sont utilisées. La première, p , représente la probabilité que vous buviez effectivement tel ou tel nombre de verres de cognac chaque soir. Ces probabilités sont obtenues grâce à une connaissance accumulée sur 10 ans par exemple. La seconde, π , représente votre capacité physique à boire tel ou tel nombre de verre de cognac. Ces possibilités ont pu être obtenues à partir d'une expérience de résistance (par exemple un défi lors d'une soirée avec des camarades de travail).

	1	2	3	4	5	6
π	1	1	0.7	0.3	0.1	0
p	0.8	0.1	0.1	0	0	0

Tableau 2.2. Distributions relatives au nombre de verre de cognac.

En "moyenne" vous buvez un verre de cognac tous les soirs, et beaucoup plus rarement 2 ou 3 verres, bien que vous soyez parfaitement capable d'en boire 2, ou même 3.

2.2.3 Axiomes généraux

Soit une distribution de possibilité $\pi: X \rightarrow [0,1]$. Elle indique le degré de possibilité de réalisation pour des événements élémentaires. X est le référentiel sur lequel est défini la distribution de possibilité, et est souvent appelé univers de discours (cadre de discernement pour les fonctions de croyance).

La distribution de possibilité doit respecter la contrainte suivante:

$$\max_x \pi(x) = 1 \tag{2.3}$$

Le paragraphe 2.2.4 examine le sens de cette contrainte.

Si la notation 2^X est introduite pour représenter l'ensemble des parties de X , une mesure de possibilité $\Pi: 2^X \rightarrow [0,1]$ est définie par:

$$\forall A \subset X, \Pi(A) = \max_{x \in A} \pi(x) \tag{2.4}$$

Par exemple cette mesure permet d'estimer la possibilité de boire sans risques 1 ou 2 verres de cognac (dans ce cas $A = \{1,2\}$).

Une mesure de nécessité N est définie par:

$$\forall A \subset X, N(A) = 1 - \Pi(A^c) \tag{2.5}$$

Le sens de cette mesure est le degré d'impossibilité de l'événement contraire.

Elle se démarque fortement des probabilités, puisqu'avec ces dernières, l'improbabilité de l'événement contraire est exactement la probabilité:

$$1 - P(A^c) = 1 - \sum_{a \in A^c} p(a) = 1 - (1 - P(A))$$

montre en effet que

$$1 - P(A^c) = P(A)$$

Par la suite, lorsque le terme mesures de possibilités (au pluriel) sera utilisé, il signifiera à la fois la mesure de possibilité Π et la mesure de nécessité N , car ces deux mesures font parties de la même théorie. La relation 2.5 permet de passer de l'une à l'autre immédiatement.

Les propriétés de ces 2 mesures sont les suivantes:

$$\Pi(\emptyset) = 0, \Pi(X) = 1, N(\emptyset) = 0, N(X) = 1$$

Ces relations énoncent simplement qu'au moins un événement se produira lors de toute expérience (car X est bien l'union de tous les événements élémentaires $x \in X$), ceci étant certain (car la nécessité de cette réalisation X est égale à 1).

D'autre part, ces relations énoncent que l'événement impossible (ou contradiction suivant le terme choisi), représenté par \emptyset , ne se produira jamais (sa possibilité étant nulle).

$$\forall A \subset X, N(A) \leq \Pi(A)$$

La nécessité d'un événement A est toujours inférieure à la possibilité du même événement, car le degré de certitude qu'il est possible d'avoir dans la réalisation d'un événement A est toujours moins importante que son degré de réalisation éventuel maximum.

Les axiomes fondamentaux des mesures de possibilités sont les suivants:

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \quad (2.6)$$

$$N(A \cap B) = \min(N(A), N(B)) \quad (2.7)$$

La relation 2.6 énonce simplement que la possibilité d'un événement union de deux événements est égale au maximum des possibilités des deux événements pris séparément, et réciproquement la relation 2.7 indique que la nécessité ou certitude en l'intersection de deux événements est égale au minimum des nécessités des deux événements pris séparément. Ce sont donc des mesures prudentes.

Il est possible de comparer ces axiomes à celui des probabilités:

$$\text{si } A \cap B = \emptyset, \text{ alors } P(A \cup B) = P(A) + P(B)$$

Les mesures de possibilités n'apparaissent donc pas du tout additives, et procèdent d'une toute autre logique que les probabilités.

Les possibilités apparaissent simples à calculer. Grâce à la relation 2.6, la relation suivante est en effet obtenue:

$$\forall A \subset X, \Pi(A) = \max_{x \in A} \pi(x) \quad (2.8)$$

Dans cette définition, et les précédentes, les sous ensembles A et B étaient classiques. Mais le premier chapitre a introduit les ensembles flous. Est-il possible de calculer les mesures de possibilités pour des ensembles flous? La réponse est oui:

$$\Pi(A) = \max_x \min(\mu_A(x), \pi(x)) \quad (2.9)$$

Ainsi il est possible de calculer le degré de possibilité que vous puissiez boire 3 **grands** verres de cognac, et rester lucide (grand étant bien sûr un concept flou).

2.2.4 Hypothèses du monde ouvert et du monde fermé

Dans la définition d'une distribution de possibilité est introduite la contrainte $\max_x \pi(x) = 1$ (autrement dit l'ensemble flou identifié à $\pi(x)$ est normalisé).

Ceci est équivalent au fait que $\prod(X) = 1$, à cause de la relation 2.8. Dire que la possibilité de l'ensemble de référence X est maximale et égale à 1 signifie simplement qu'au moins un des événements x a un degré de possibilité égal à 1, c'est à dire qu'au moins un des événements x se réalisera.

Si cette contrainte n'était pas respectée, c'est à dire si $\max_x \pi(x) < 1$ alors il pourrait se produire un événement "impossible" (s'il est possible de lui donner un nom), encore représenté par l'ensemble \emptyset . En effet sa mesure de nécessité serait strictement supérieure à 0, puisque $N(\emptyset) = 1 - \prod(\emptyset^c) = 1 - \prod(X) > 0$. Il en serait de même pour sa mesure de possibilité $\prod(\emptyset)$ (car la mesure de possibilité est toujours supérieure à la mesure de nécessité).

Or par définition l'ensemble X est l'ensemble de tous les événements x possibles. Ceci serait en contradiction avec le fait que l'événement impossible pourrait se réaliser.

La contrainte $\max_x \pi(x) = 1$ est donc connue sous le nom d'hypothèse du monde fermé, et résume simplement que le fait que l'ensemble X contient tous les événements possibles, ou encore qu'ils ont tous été énumérés dans l'ensemble X [Dub 1985].

Avant de définir le monde ouvert, un petit détour est nécessaire. Le paragraphe 2.3 aborde les mécanismes de fusion de sources d'informations. En quelques mots, la fusion est appliquée lorsque plusieurs sources d'informations sont rassemblées, *fusionnées*, pour générer une connaissance globale.

Souvent dans les mécanismes de fusion, sont appliqués des mécanismes de normalisation: après une fusion, la distribution résultante π_{res} est sous normalisée, c'est à dire que $\max_x \pi_{res}(x) < 1$. La renormalisation est un procédé qui permet de transformer le résultat π_{res} en une distribution normalisée.

Plutôt que de normaliser, et de se placer dans l'hypothèse du monde fermé, il est possible de se placer dans l'hypothèse du monde ouvert. L'exemple suivant permet de mieux appréhender cette hypothèse:

Une voiture a commis un accident, et le conducteur fautif s'est enfuit. Les policiers arrivés par la suite interrogent différentes personnes, chacune constituant une source d'information, pour connaître la marque, le type, la couleur de la voiture.

Or justement les deux seuls témoins ne s'entendent pas sur la couleur de la voiture. A dit 'bleu', B dit 'rouge', parmi un ensemble de possibilités Couleur={bleu, rouge, vert, noir, blanc}.

Ces deux sources d'informations sont en complète contradiction. A cause de l'hypothèse du monde fermé, la fusion de ces deux sources est complexe, et comme résultats pourraient être obtenus soit π_1 définie par $\pi_1(\{x\}) = 1 \forall x \in \text{Couleur}$, soit π_2 définie par $\pi_2(\{\text{bleu}\}) = \pi_2(\{\text{rouge}\}) = 1$.

Si l'hypothèse du monde ouvert était choisie [Dubois et Prade, 1988], lors de la fusion, il y aurait une sorte de signal d'alerte, dû au fait que les deux seules sources d'information seraient en complète opposition, et d'autres mécanismes de fusion pourraient être envisagés, en modifiant l'ensemble Couleur.

En effet les deux témoins sont peut être daltoniens, et la voiture est peut être simplement 'violette'. Dans ce cas, sur l'ensemble Couleur' = Couleur \cup {violet}, la fusion serait très naturelle, les deux sources étant en parfait accord, et ayant toutes les deux raison.

En résumé l'hypothèse du monde fermé est toujours valide en statique, lorsque des distributions sont simplement observées. Elle signifie simplement que le référentiel est complet. Par contre, en dynamique, lorsque les informations sont fusionnées, l'hypothèse du monde ouvert peut être choisie. C'est une autre manière d'interpréter les conflits entre sources d'informations, en acceptant le fait que le référentiel, noté ensemble X, soit peut être incomplet. Dans ce cas, d'autres méthodes de fusion sont envisagées.

Dans l'exemple de la couleur des voitures, d'autres couleurs que celles référencées dans l'ensemble Couleur, auraient été envisagées, pour vérifier si les deux sources n'étaient pas d'accord en fin de compte.

Ce problème du monde ouvert face au monde fermé est universel, et est aussi abordé dans la théorie de Dempster Shafer. Dans la suite de ce chapitre, et donc dans la partie fusion, l'hypothèse du monde fermé est toujours choisie.

2.2.5 Sens des mesures de possibilités

A coté des axiomes et propriétés définissant les possibilités au paragraphe 2.2.3 a été donnée une interprétation plus intuitive des possibilités comme étant des degrés mesurant une capacité physique des événements à se réaliser.

D'autres interprétations peuvent en être tirées, au moyen de fonctions mesurant telle ou telle quantité.

Ainsi dans la théorie des probabilités a été définie l'entropie de Shannon:

$$H(p) = - \sum_{x \in X} p(x) \log_2(p(x)) \quad (2.10)$$

Cette fonction mesure la quantité d'information d'une distribution de probabilités.

L'exemple qui suit montre ce qu'est la quantité d'information contenue dans une distribution p.

Le sort du monde dépend du prochain résultat du lancer d'une pièce: pile, le paradis, face l'enfer. Deux pièces sont considérées. A est telle que $p_A(\{\text{pile}\}) = p_A(\{\text{face}\}) = 0.5$ et B est telle que $p_B(\{\text{face}\})=1$. B est une pièce truquée.

Les deux mesures d'entropie sont les suivantes: $H(p_A) = \log_2 2$ et $H(p_B) = 0$.

L'entropie de Shannon pour la première pièce A est maximale, en étant égale au logarithme du nombre d'états de la pièce (ici deux états seulement).

En fait la quantité d'information fournie par le résultat de la pièce A est très importante, car rien ne permet de dire si ce sera le paradis ou l'enfer sur terre avant l'expérience.

Au contraire l'expérience pour le cas de la pièce B est inutile, et n'apportera aucune information, puisque étant donné la distribution de probabilité le sort de la terre est déjà réglé.

Pour la théorie des possibilités, des fonctions du même type que l'entropie de Shannon ont aussi été développées.

En fait deux mesures ont été définies, et ce dans la théorie de Dempster Shafer [Klir, 1987, 1994]. Ces deux mesures ont été définies pour les fonctions de croyance, et n'ont pu être appliquées aux possibilités que parce qu'il existe des relations de passage entre les possibilités et les fonctions de croyance. Il n'est pas possible d'en comprendre clairement la nature sans passer par la théorie de Dempster Shafer, aussi seuls les résultats finaux seront donnés.

La première mesure est une généralisation de l'entropie de Shannon aux fonctions de croyance. C'est une mesure de la contradiction interne à une source d'information. Par exemple dans l'exemple de la couleur de la voiture au chapitre 2.2.4, une possibilité aurait été $\pi_2(\{\text{bleu}\}) = \pi_2(\{\text{rouge}\}) = 1$, ce qui était plutôt contradictoire.

Le résultat de cette mesure de contradiction pour une distribution de possibilité π est :

$$H(\pi) = - \sum_{i=1}^{\text{Card}(X)} (\pi(x_i) - \pi(x_{i+1})) \log_2 \left[1 - i \sum_{j=i+1}^{\text{Card}(X)} \frac{\pi(x_j)}{(j-1)j} \right] \quad (2.11)$$

avec par hypothèse une permutation sur les indices tels que $\pi(x_1) = 1 \geq \pi(x_2) \geq \dots \geq \pi(x_{\text{Card}(X)+1}) = 0$.

Une deuxième mesure quantifie la non précision d'une source d'information:

$$S(\pi) = \sum_{i=2}^{\text{Card}(X)} \pi(x_i) \log_2 \frac{i}{i-1} \quad (2.12)$$

Comme l'entropie de Shannon, ces deux mesures varient entre 0 et $\log_2 \text{Card}(X)$, où $\text{Card}(X)$ représente le nombre d'états différents du référentiel.

Une conclusion importante sur ces deux mesures est que lorsque le nombre d'états du système augmente, alors la mesure de contradiction devient beaucoup plus petite que la mesure de non précision: $H(\pi) \ll S(\pi)$ lorsque $\text{Cardinal}(X) \rightarrow \infty$.

Autrement dit les mesures de possibilités quantifient mal les conflits, et sont beaucoup plus dédiées à la représentation de l'imprécision des informations.

2.2.6 Représentations usuelles

Il existe certaines formes canoniques de l'imprécision modélisées par les mesures de possibilités. Ce paragraphe présente les modèles de données qui sont parfaites, imprécises, ou en même temps imprécises et incertaines, pour des variables numériques. Ils peuvent être étendus aux autres types de données sans difficultés.

2.2.6.1 Information précise et certaine

La figure 2.3 montre un exemple d'information précise et certaine:

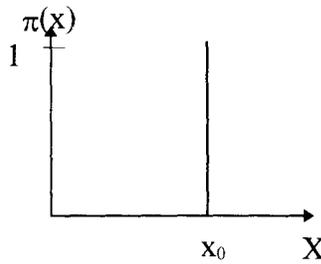


Fig. 2.3. Information précise et certaine.

La représentation de la certitude que l'événement x_0 se produira est modélisée par une distribution π telle que $\pi(x_0)=1$ et $\pi(x)=0 \forall x \in X - \{x_0\}$.

En effet pour une telle distribution, $\prod(\{x_0\})=N(\{x_0\})=1$. De même pour une telle information, la mesure de contradiction $H(\pi)$ égale 0 et la mesure de non spécificité $S(\pi)$ égale 0 aussi. Le cœur d'une information précise et certaine est confondu avec son support, les deux étant réduits au même singleton. L'ensemble flou associé à une telle distribution est appelé un singleton.

2.2.6.2. Information imprécise

Comment représenter l'imprécision d'une mesure fournie par un capteur qui donne une information x_0 ? L'approche consiste à assigner aux événements $x_0+\varepsilon$ (avec ε petit), proches de l'événement principal x_0 , un degré de possibilité non nul. Plus l'événement $x_0+\varepsilon$ est proche de x_0 (c'est à dire plus ε est petit), plus le degré de possibilité est proche de 1, et décroît à mesure que ε croît.

L'approche la plus utilisée en théorie des possibilités et une approche linéaire, comme un exemple le montre à la figure 2.4.

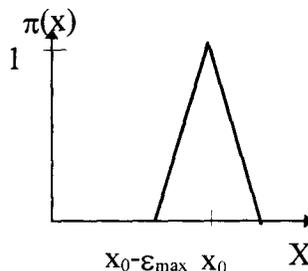


Figure 2.4. Information imprécise centrée en x_0 .

Le seul paramètre dans cette approche est le facteur ε_{\max} au delà duquel le degré de possibilité atteint 0 (autrement dit c'est la base du triangle).

Pourtant il serait possible d'envisager d'autres modélisations de l'imprécision, suivant le capteur, notamment si des informations statistiques quant à la dispersion des informations sont disponibles (pourquoi pas des courbes gaussiennes?).

Le cœur d'une information imprécise est un singleton, qui est contenue dans le support, en général centré sur le cœur.

2.2.6.3 Information incertaine

Dans la théorie des possibilités, lorsque le terme "incertaine" est utilisé à propos d'une information, il est compris comme un degré de la confiance qui est accordée à la donnée, comme un niveau de fiabilité.

Pour modéliser la confiance, un paramètre supplémentaire est requis pour caractériser la donnée. En général ce paramètre de confiance α appartient à $[0,1]$. Une information certaine, comprise comme sûre et fiable, a un paramètre α égal à 1, une information absolument non fiable a un paramètre égal à 0.

Etant donné une information, modélisée par une distribution de possibilité, et un paramètre de confiance α , une transformation est appliquée à π de telle sorte que la confiance α soit directement incluse dans une nouvelle information π' qui résume alors, et π , et α .

Le principe de la transformation est le suivant: dire qu'une information est plus ou moins sûre revient à relativiser les informations fournies de telle sorte qu'aucun événement ne peut plus être impossible. Ainsi tout événement x aura un degré de possibilité non nul, sauf si l'information fournie est certaine.

Ce paragraphe donne les deux types de transformations qui ont été définies, et donne un exemple d'application à une information imprécise.

Les deux types de transformations sont d'une part une approche logique, d'autre part une approche linéaire [Dubois et Prade, 1992b].

La relation pour l'approche logique est:

$$T_{DP}(\pi, \alpha) = \max(\pi, 1 - \alpha) \tag{2.13}$$

La relation pour l'approche linéaire est:

$$T_{Yag}(\pi, \alpha) = \alpha\pi + 1 - \alpha \tag{2.14}$$

Dans les deux cas l'indice est relatif aux initiales des chercheurs associés aux transformations.

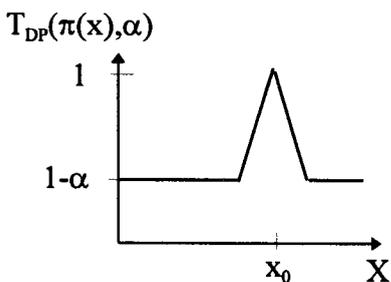


Fig. 2.5. Information imprécise et d'incertitude α , approche logique.

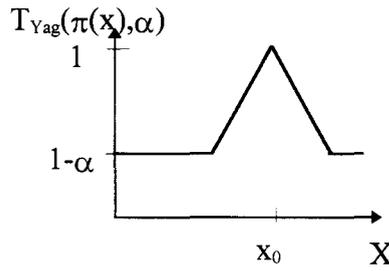


Fig. 2.6. Information imprécise et d'incertitude α , approche linéaire.

Une hypothèse implicite sur laquelle sont basées les deux approches est que la modélisation de variation de confiance en fonction du paramètre α est aussi linéaire. Ainsi une information de degré de confiance α égal à 1 est considérée comme étant deux fois plus fiable qu'une information de confiance 0.5.

Or rien ne dit que les humains modélisent subjectivement ainsi leur sentiment de confiance. Il pourrait très bien être envisagé d'autres transformations qui privilégient (ou qui altèrent moins) les données fiables. Ainsi la transformation suivante, toujours dans l'approche linéaire, renforce le sentiment de méfiance:

$$T'_{Yag}(\pi, \alpha) = \alpha^2 \pi + 1 - \alpha^2 \quad (2.15)$$

Il est même possible des transformations encore plus différentes en changeant simplement de t-conorme:

$$T(\pi, \alpha) = \min(\pi + 1 - \alpha, 1) \quad (2.16)$$

Dans les deux cas, les transformations ont le même comportement limite qu'avec les relations 2.13 et 2.14.

Il faut enfin noter que le support d'une information incertaine est l'ensemble du référentiel.

2.2.6.4 Ignorance absolue

Il existe une forme d'information très particulière appelée ignorance absolue ou encore complète.

Elle est représentée par une distribution π définie par $\pi(x) = 1 \quad \forall x \in X$.

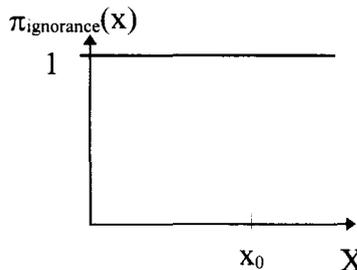


Fig. 2.7. Ignorance complète.

Tous les événements sont également possibles, tous ont une nécessité nulle. En fait cette distribution n'apporte aucune information. Elle est équivalente en termes de possibilités à une distribution de probabilité uniforme.

2.2.7 Construction des degrés de possibilités

L'estimation des degrés de possibilités de réalisation d'événements est un problème complexe. Elle peut parfois relever du domaine subjectif lorsque des experts humains sont interrogés, parfois elle peut être déduite de certaines lois physiques ou mathématiques. Parfois elle peut aussi être liée à une certaine connaissance statistique des variables mises en jeu.

Ce paragraphe revient sur les rapports entre possibilités et probabilités, rapports simplement évoqués au paragraphe 2.2.2.

2.2.7.1 Compatibilité entre distributions de probabilités et de possibilités

Les possibilités mesurent la capacité à se réaliser de certains événements. Les probabilités mesurent des connaissances sur la réalisation de certains événements, connaissances liées à la répétition de certaines expériences.

Logiquement un événement ne pourrait réellement se réaliser que si son degré de possibilité n'était pas nul. En fait il existe un principe entre degré de probabilité et degré de possibilité qui lie les valeurs de ces deux degrés.

Le principe de compatibilité ou de cohérence s'énonce ainsi [Delgado et Moral, 1987]:

$$\forall x \in X, p(x) \leq \pi(x) \quad (2.17)$$

Il énonce qu'un événement ne peut pas avoir plus de chances probabilistes de se réaliser qu'il n'a de degrés de possibilités de se réaliser.

Il est alors possible de comparer deux types d'informations, représentées par une distribution de probabilité et une distribution de possibilité.

Dans ces conditions les degrés de possibilités sont compris comme une connaissance a priori que vient confirmer ou infirmer une expérience. Ceci n'est valable bien entendu que pour les variables qui peuvent être soumises à des expériences.

Des fonctions ont été définies pour vérifier alors si la connaissance a priori (c'est à dire les possibilités) est compatible avec les mesures.

La plus simple est binaire:

$$C(p, \pi) = \begin{cases} 1 & \text{si } p(x) \leq \pi(x) \\ 0 & \text{sin on} \end{cases} \quad (2.18)$$

Zadeh a proposé

$$C(p, \pi) = \sum_{x \in X} p(x)\pi(x) \quad (2.19)$$

et Sugeno

$$C(p, \pi) = \sup_{A \subset X} (\cap (\Pi(A), P(A))) \quad (2.20)$$

avec \cap une t-norme usuelle, par exemple un minimum.

2.2.7.2 Construction de degrés de probabilité à partir d'histogrammes

Lorsqu'une information a priori est disponible sous la forme d'une distribution de probabilités, il est possible d'en construire une distribution de possibilités, puisque les deux notions sont liées, comme l'énonce le principe de compatibilité.

Afin de construire effectivement les degrés de possibilités, il faut cependant faire d'autres hypothèses sur les liens entre les deux informations.

Les relations suivantes, obtenues par Dubois et Prade [Dubois 1983], sont fondées sur l'hypothèse qu'une variation de probabilité entre deux événements doit être comprise comme un degré de certitude supplémentaire en faveur de la plus probable. En travaillant sur les mesures de nécessités, puis en les traduisant en distributions de possibilités, le résultat final apparaît finalement sous la forme suivante:

$$\pi(x_i) = ip(x_i) + \sum_{j=i+1}^{\text{Card}(X)} p(x_j) \quad (2.21)$$

avec par convention $\sum_{j=\text{Card}(X)+1}^{\text{Card}(X)} p(x_j) = 0$ et les x_i ordonnés de telle sorte que $p(x_1) \geq \dots \geq p(x_{\text{Card}(X)})$.

2.3 Fusion de données

La fusion de données est un aspect très important de la théorie des possibilités. Bien qu'étant conçue pour traiter des informations imparfaites, il est toujours plus agréable d'utiliser des données précises et fiables.

Ainsi lorsqu'un homme affronte un problème complexe basé sur des données incertaines, il cherche toujours à agréger plusieurs sources différentes d'information afin d'augmenter la fiabilité des données, et de réduire l'imprécision des mesures. Les méthodes du juge d'instruction en sont un bon exemple: non seulement il cherche à obtenir le maximum de témoignages différents, mais en plus il cherche aussi le maximum de preuves matérielles, toutes ces informations concourant à ce qu'il forge son intime conviction. Ainsi pour obtenir un bon portrait robot, telle personne aura remarqué plus particulièrement la forme du menton, une autre la forme des oreilles, une autre encore la couleur des yeux.

Pris séparément, chacun des témoignages ne donne qu'une vue très restreinte du coupable éventuel, mais en fusionnant l'ensemble des témoignages, un portrait robot relativement clair peut être construit. De même, plus le nombre de témoignages concordent pour un élément précis, plus les policiers sont contents.

La théorie des possibilités a la capacité de pouvoir effectuer les mêmes opérations de fusion d'informations. En sus de l'hypothèse du monde fermé, une autre hypothèse est nécessaire avant de présenter les mécanismes de fusion.

Pour qu'il y ait fusion, il faut plusieurs sources d'informations. Il faut même, s'il existe effectivement plusieurs sources d'informations, que ces sources soient distinctes. La Cour des Comptes avait ainsi relevé plusieurs anomalies dans la gestion de l'ARC. Une des anomalies concerne plus particulièrement la fusion de donnée. En effet l'ARC donnait de l'argent à différent laboratoires pour soutenir leur recherche. Il fallait donc évaluer ces laboratoires pour savoir qui méritait de l'argent, et qui n'en méritait pas. L'ARC avait donc mis en place des

systèmes d'évaluation faisant appel à des experts, pour noter chaque laboratoire. La Cour des Comptes a alors relevé que le professeur A devait évaluer le laboratoire du professeur B, et que réciproquement, le professeur B devait évaluer le laboratoire du professeur A.

En fait avec les problèmes de fusion il faut toujours analyser cette question d'indépendance entre les informations. En effet certains opérateurs éliminent les sources redondantes (min), d'autres non (le produit). Si deux sources d'informations sont fusionnées, alors qu'elles fournissent des données pratiquement identiques, il est logique que le sentiment de confiance ou de fiabilité dans le résultat augmente, sauf si justement les données sont liées. C'est le problème des sources d'informations dépendantes: elles n'apportent pas plus d'information, mais provoquent une dérive dans le sentiment de certitude dans leurs informations. Suivant le type de comportement retenu (renforcement ou non), il faudra choisir les opérateurs en fonction de l'indépendance. Le principal problème reste actuellement la détection des sources liées. Ce problème est le plus souvent rencontré lorsque le facteur humain est impliqué [Dubois et Prade, 1992b].

Dans la suite l'hypothèse d'indépendance des sources est toujours admise.

Une dernière remarque sera faite avant de présenter les méthodes de fusion: il ne faut pas confondre la fusion avec la révision (encore appelée fusion dissymétriques). La fusion est un processus qui agrège les informations en parallèle, tandis que la révision est un processus par lequel une source d'information est modifiée dans le temps par l'apport d'autres connaissances certaines. La révision implique des approches semblables aux probabilités conditionnelles, et n'est pas traitée dans ce paragraphe. Le but de ce paragraphe est de présenter la plupart des outils actuels de fusion symétrique de données dans la théorie des possibilités. Quatre grandes approches seront d'abord présentées au paragraphe 2.3.1 puis quelques opérateurs ayant été proposés pour combler certains défauts des méthodes classiques seront montrés aux paragraphes 2.3.2, 2.3.3 et 2.3.4.

Afin d'exposer clairement les résultats, la plupart des opérateurs seront utilisés sur deux exemples.

Le premier est relatif à un référentiel X continu: $X=[-0.75,0.75]$. Les distributions de possibilités à fusionner sont deux trapèzes. Afin de donner le maximum d'information, les résultats seront donnés en 3D avec un paramètre supplémentaire, la contradiction entre les deux trapèzes. Pour cela l'écartement entre les deux trapèzes variera. La figure 2.8 montre trois cas particuliers de fusion.

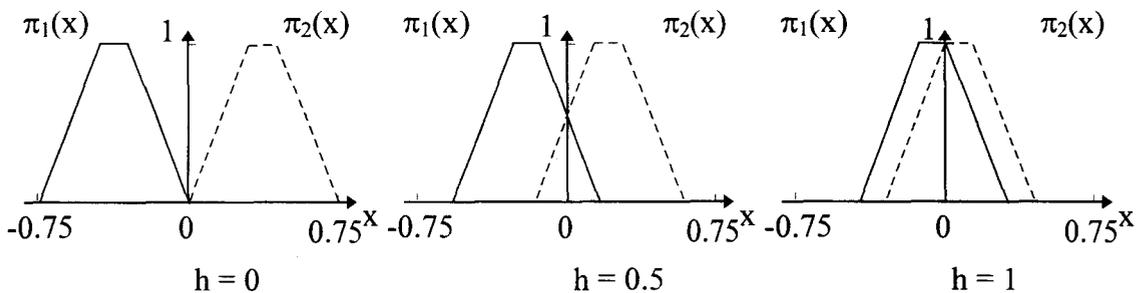


Fig. 2.8. Trois type de configuration pour le premier exemple. Le paramètre h mesure la non contradiction, et varie entre 0 et 1: $h = 0$ représente ainsi une contradiction absolue.

Dans le cas d'une contradiction totale, h (pour hauteur de l'intersection entre les deux trapèzes) vaut 0, et les deux trapèzes se touchent juste à la limite. Puis, à mesure que la contradiction diminue, les deux trapèzes sont rapprochés de sorte que l'intersection augmente petit à petit. Lorsque le paramètre h vaut 1, c'est à dire lorsque l'intersection a une hauteur de 1, les deux trapèzes ne sont pas nécessairement confondus, comme le montre la figure 2.8.

Le second exemple est relatif à la détermination de la couleur d'une voiture, par deux sources d'informations. Le référentiel X est dans ce cas l'ensemble discret {rouge, bleu, jaune, vert...}.

2.3.1 Définitions et opérateurs classiques

Dans ce paragraphe les principes de fusion sont abordés avec les opérateurs conjonctif, disjonctif, adaptatif et linéaire, qui représentent les principales méthodes de fusion. Les notions de contradiction et de normalisation sont également abordées.

2.3.1.1 Intersection: la fusion conjonctive

Cette règle est la plus utilisée, et est à la base de toutes les autres extensions.

Son principe est l'intersection: si une source dit que x (recherché) appartient à A et qu'une autre source dit que x appartient à B , alors il est raisonnable, si ces sources sont fiables, de conclure que x appartient à la fois à A et à B .

Ceci se résume par l'expression $(x \in A \text{ et } x \in B) \Rightarrow x \in A \cap B$. Cette méthode est encore appelée fusion conjonctive.

La première étape consiste à calculer l'intersection des distributions de possibilité, tout comme l'intersection d'ensembles flous, par un opérateur de Zadeh:

$$\begin{aligned} \forall x \in X \\ \pi_{\cap}^*(x) = \cap(\pi_1(x), \dots, \pi_n(x)) = \min_i \pi_i(x) \end{aligned} \quad (2.22)$$

Est introduite alors la fonction h estimant la hauteur d'une distribution $\pi(x)$:

$$h(\pi(x)) = \max_x \pi(x) \quad (2.23)$$

(Utilisée sans argument, la notation h fait simplement référence à la hauteur d'une distribution de possibilité résultant de l'intersection d'autres sources: $h = \max_{x \in X} \pi_{\cap}^*(x)$).

La hauteur d'une distribution est normalement toujours égale à 1, à cause de l'hypothèse du monde fermé. Elle n'est en fait utilisée que comme paramètre intermédiaire lors de processus de fusion. En effet la relation 2.22 définissant l'intersection de possibilités ne donne que très rarement un ensemble flou normalisé, comme le montre la figure 1.5. Il faut alors une seconde étape, dite de normalisation.

La seconde étape consiste donc à normaliser le résultat obtenu par la relation 2.22.

Est d'abord défini un opérateur de normalisation No appliqué à une fonction $f: X \rightarrow [0,1]$

$$No(f(x)) = \frac{f(x)}{h(f(x))} = \frac{f(x)}{\max_x (f(x))} \quad (2.24)$$

Finalement est définie la forme finale de la règle de fusion conjonctive:

$$\forall x \in X$$

$$\pi_{\cap}(x) = \text{No}(\pi_{\cap}^*(x)) = \frac{\pi_{\cap}^*(x)}{h} = \frac{\min_i \pi_i(x)}{\max_x \min_i \pi_i(x)} \quad (2.25)$$

La hauteur d'une distribution de possibilité est utilisée comme indicateur de la pertinence de l'opérateur de fusion choisi, ou comme indicateur de la qualité du résultat. En l'occurrence pour l'intersection, si le processus de normalisation est justifié lorsque h est proche de 1, par exemple h égalant 0.7, ce n'est plus du tout vrai lorsque h est proche de 0. En effet une hauteur proche de zéro signifie que les sources sont en forte contradiction, puisqu'elles n'ont presque pas de point commun.

Or par hypothèse d'application de l'intersection, les sources sont fiables, donc doivent fournir les mêmes informations, ou presque. C'est ainsi que la relation 2.25 n'est même pas définie dans le cas où h vaut 0.

L'intersection définie par la relation 2.25 est donc un opérateur discontinu.

La figure 2.9 montre le résultat avec l'exemple de fusion des trapèzes.

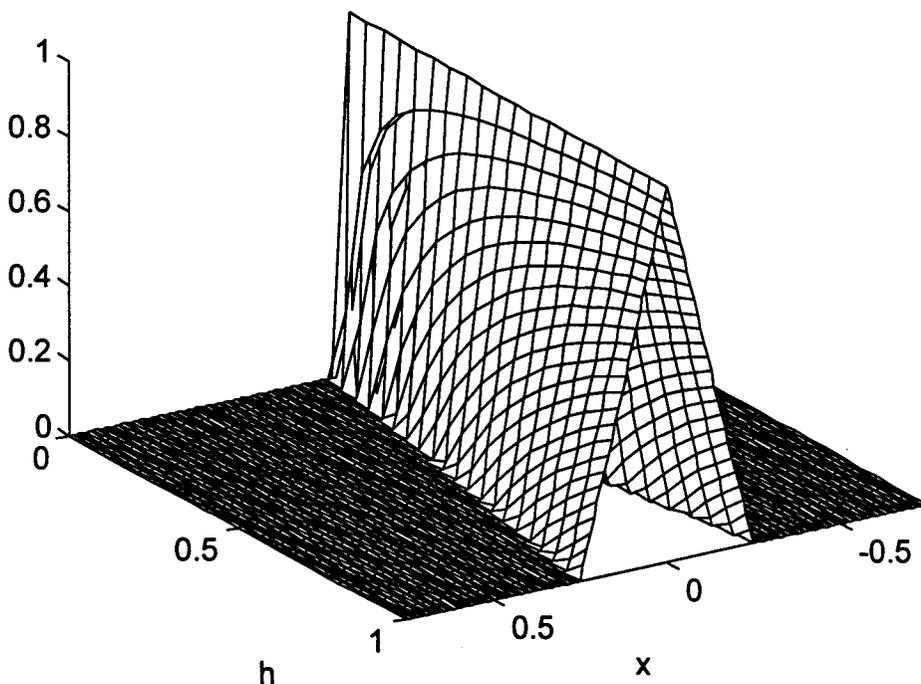


Fig. 2.9. Fusion conjonctive des sources π_1 et π_2 .

Lorsque $h = 1$, l'intersection est parfaitement obtenue, et à mesure que la contradiction augmente, la surface se rétrécit de plus en plus, avec un maximum égal à 1 en $x = 0$, qui à la limite devient une impulsion. Rien n'est tracé pour $h = 0$.

L'exemple des couleurs insiste sur ce problème:

Les personnes A et B apportent des informations sous forme de distribution π_A et π_B :

$\pi_A(\text{rouge}) = 1$, $\pi_A(\text{jaune}) = 10^{-50}$ et zéro pour les autres couleurs.

$\pi_B(\text{vert}) = 1$, $\pi_B(\text{jaune}) = 10^{-50}$ et zéro pour les autres couleurs.

Avant normalisation, le résultat est $\pi(\text{jaune}) = 10^{-50}$ et zéro pour les autres couleurs. Après la normalisation le résultat devient $\pi(\text{jaune}) = 1$. Ainsi les deux sources d'informations, qui étaient en désaccord total, sauf à dire que la couleur jaune était un infiniment petit possible, ont, par l'application de la règle conjonctive, donné un résultat qui affirme que la seule couleur possible était le jaune!

En résumé la règle de fusion conjonctive est la plus utilisée. Son application suppose que les sources soient fiables, et dans le cas contraire elle a un comportement à la limite discontinu.

Afin de remédier à ce comportement un chercheur a proposé un autre opérateur de normalisation No', défini par:

$$\text{No}'(f(x)) = f(x) + 1 - h(f(x)) \quad (2.26)$$

L'opérateur No' fait un rééchelonnement des degrés de possibilités. L'opérateur No' ajoute aux degrés de possibilité la quantité de contradiction.

Avec l'intersection, le résultat est finalement

$$\pi_Y(x) = \pi_{\cap}^*(x) + 1 - h = \min_i \pi_i(x) + 1 - \max_x \inf_i \pi_i(x) \quad (2.27)$$

L'indice Y se réfère à Yager qui a proposé cette normalisation. Bien qu'étant parfaitement continu, cet opérateur est peu utilisé, car il est beaucoup moins informatif et précis que le précédent, puisque le support du résultat est la totalité du référentiel.

La figure 2.10 montre cet opérateur appliqué aux deux sources d'information.

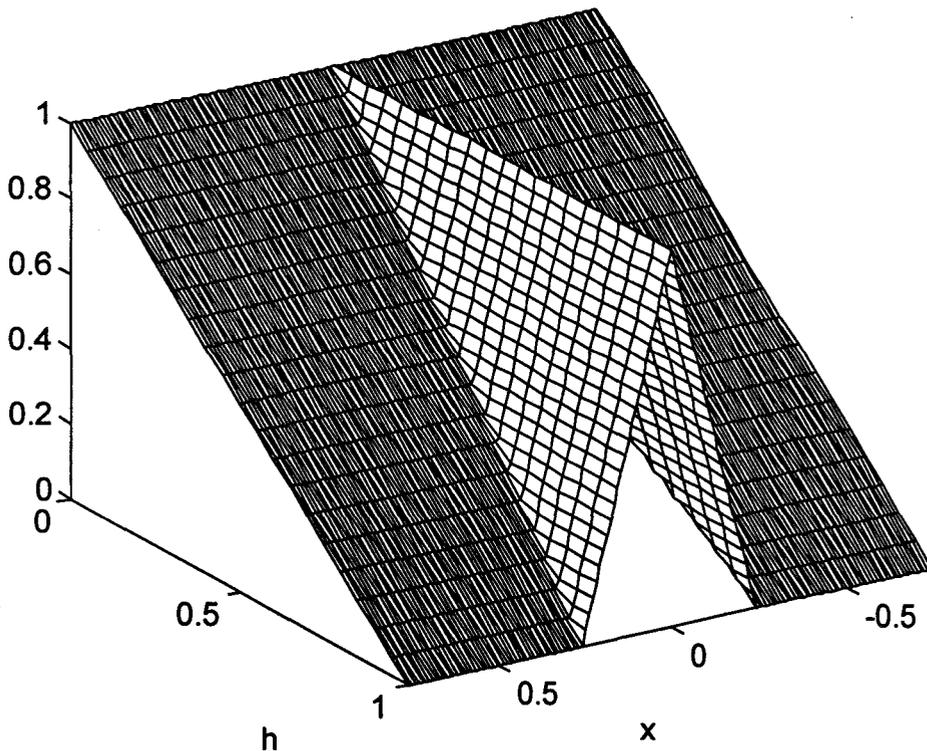


Fig. 2.10. Conjonction avec la normalisation de Yager.

Dans le cas où h est proche de 0, tous les événements sont considérés également possibles.

2.3.1.2 Union: la fusion disjonctive

L'union se définit par rapport à l'intersection. En fait elle s'applique lorsque les sources ne sont pas fiables, à part une, mais sans savoir laquelle est bonne.

L'intersection ne serait pas une approche adéquate, puisque la bonne information serait polluée par les autres, avec tous les problèmes de contradiction et de discontinuité.

Aussi, afin de conserver le maximum de données pertinentes, une attitude prudente est de garder, d'englober, toutes les informations. Ceci est réalisé par une simple union:

$$\forall x \in X$$

$$\pi_{\cup}(x) = \cup(\pi_1(x), \dots, \pi_n(x)) = \max_i \pi_i(x) \quad (2.28)$$

Cette méthode de fusion est appelée fusion disjonctive. Elle donne bien sur des informations beaucoup moins précises que l'intersection, mais dans le cas d'informations peu sûres, c'est le seul mode adéquat. La figure 2.11 en donne les résultats obtenus par la relation 2.28 (le résultat est directement normalisé).

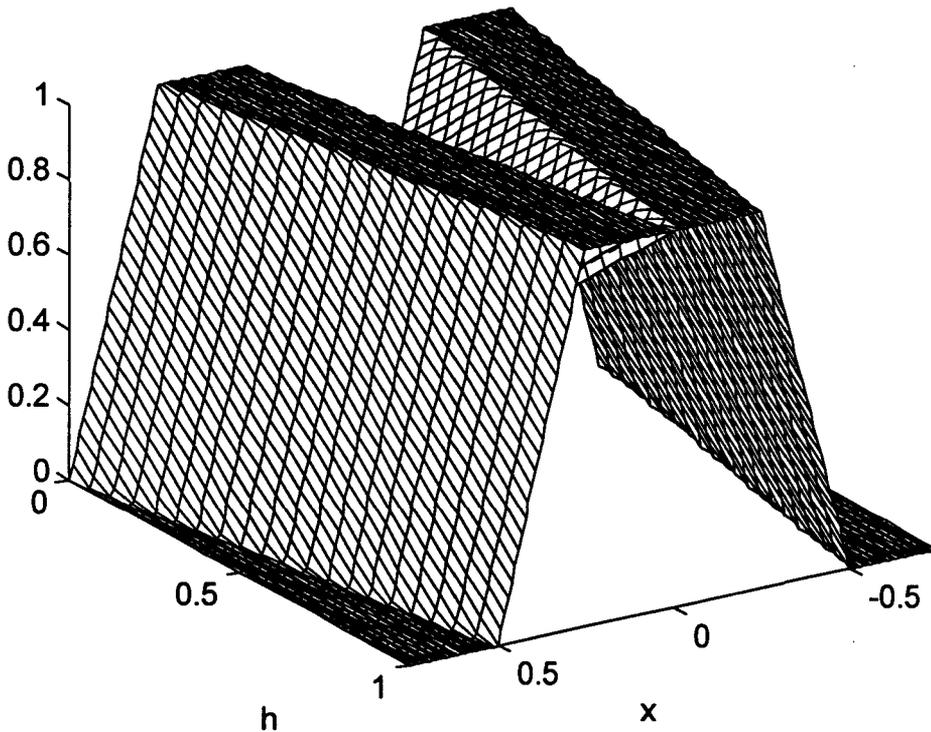


Fig. 2.11. Fusion disjonctive.

L'exemple de la couleur donne le résultat de la règle disjonctive appliquée à des sources contradictoires:

A dit que $\pi_A(\text{rouge}) = 1$, $\pi_A(\text{jaune}) = 10^{-50}$

B dit que $\pi_B(\text{vert}) = 1$, $\pi_B(\text{jaune}) = 10^{-50}$

Avec la règle disjonctive, le résultat est:

$\pi(\text{rouge}) = 1$, $\pi(\text{vert}) = 1$, $\pi(\text{jaune}) = 10^{-50}$.

En fait deux couleurs restent possibles, rouge ou vert, et le jaune ne devient pas la couleur la plus probable.

2.3.1.3 Fusion adaptative de Dubois et Prade

Les deux modes de fusion, conjonctive et disjonctive, sont les deux principaux modes de fusion. Ils ont un comportement logique et naturel, mais leurs comportements ne peuvent s'adapter à la qualité des informations. Si les informations sont fiables, alors c'est l'intersection qu'il faut choisir, mais il faut être certain de la qualité des données. Sinon l'intersection donne des résultats aberrants. Au contraire l'union peut toujours être appliquée (sauf dans le cas où toutes les informations sont mauvaises, mais dans ce cas il n'y a pas grand chose à faire). Mais si les données sont de bonne qualité, elles seront gâchées, car l'union ne mettra pas clairement en évidence l'accord entre les sources comme le fait l'intersection.

Aussi certains opérateurs ont été conçus de telle sorte qu'ils s'adaptent d'eux mêmes à la qualité des informations.

Tout d'abord il faut mesurer la qualité des informations. Ceci est réalisé en calculant la hauteur h de l'intersection entre les sources. Si les sources sont en accord, cette hauteur est proche de 1, et si les sources sont en profond désaccord, h sera égal à 0.

Donc si h est proche de 1, une intersection sera appliquée, et si h est proche de 0, ce sera une union.

L'opérateur le plus connu ayant ce type de comportement est un opérateur conçu par Dubois et Prade [Dubois et Prade, 1994a, 1994b]:

$$\pi_{DP}(x) = \max(\pi_{\cap}(x), \min(1 - h, \pi_{\cup}(x))) \quad (2.29)$$

En fait dans cet opérateur l'intersection π_{\cap} et l'union π_{\cup} sont toujours calculées. Le premier terme \max fait en fait une union de ces deux modes de fusion qui s'interprète comme une certitude que soit l'intersection soit l'union était l'opérateur le plus approprié. C'est un comportement prudent.

Pour ne pas accorder trop d'importance à l'union, le terme $1-h$ minore son importance par le terme \min : l'union n'intervient réellement qu'en cas de désaccord, lorsque h est proche de 0.

La figure 2.12 montre le résultat sur l'exemple numérique (2.29 est normalisé par construction).

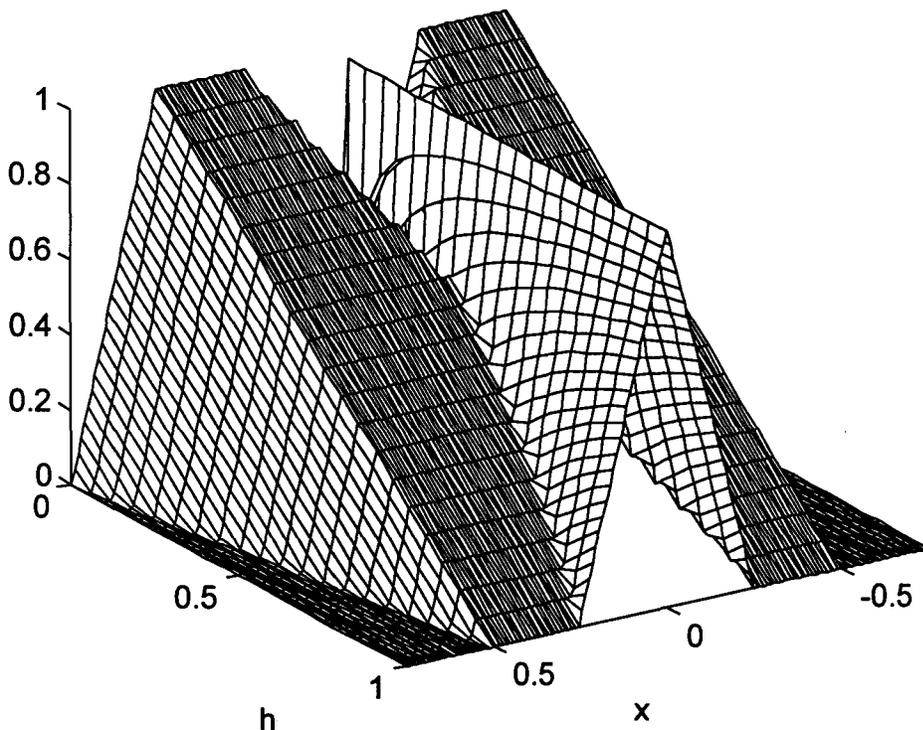


Fig. 2.12. Fusion adaptative de Dubois et Prade.

Aux deux cas extrêmes $h=1$ et $h=0$, l'intersection et l'union des sources sont bien retrouvées. Pourtant l'intersection, dans le cas présent $x=0$ en est le maximum, est systématiquement renforcée. A cause de l'utilisation de la fusion conjonctive, cette règle est à nouveau discontinue: pour $h=0$, $\pi(x=0)=0$, mais pour $h=\epsilon$ avec $\epsilon \ll 1$, $\pi(x=0)=1$.

Ceci peut se vérifier avec l'exemple de la couleur:

A dit que $\pi_A(\text{rouge}) = 1$, $\pi_A(\text{jaune}) = 10^{-50}$ (0 pour les autres)

B dit que $\pi_B(\text{vert})=1$, $\pi_B(\text{jaune}) = 10^{-50}$ (0 pour les autres)

Les deux experts sont d'accord pour dire que jaune a un très très faible degré de possibilité. Pourtant avec la règle adaptative de Dubois et Prade, le résultat est:

$\pi(\text{rouge}) = 1-10^{-50}$, $\pi(\text{vert}) = 1-10^{-50}$, $\pi(\text{jaune}) = 1$ (0 pour les autres).

Bien que les trois couleurs soient également possibles, la couleur jaune a quand même un léger avantage, et est finalement la couleur qui a le degré de possibilité le plus élevé.

2.3.1.4 Fusion linéaire.

L'opérateur présenté maintenant n'a absolument aucun lien avec les précédents, qui avaient notamment des comportements logiques de type intersection ou union.

Il vient en fait du domaine des probabilités, où c'est le seul moyen de fusionner des distributions de probabilités.

C'est une somme convexe des différentes sources de données: à chaque source π_i est associé un paramètre $\lambda_i \in [0,1]$. Ce paramètre représente le poids relatif d'une source par rapport aux autres.

L'ensemble des paramètres λ_i est lié par la contrainte:

$$\sum_i \lambda_i = 1 \quad (2.30)$$

La règle est alors

$$\pi_i(x) = \text{No}(\sum \lambda_i \pi_i(x)) \quad (2.31)$$

Cet opérateur est le premier à faire intervenir des paramètres extérieurs aux informations modélisées par des possibilités. Il est donc d'une famille totalement différente des précédents, sans compter sa différence fondamentale avec les opérateurs logiques.

Bien entendu il faut auparavant évaluer les paramètres λ_i , mais leur connaissance permet de rejeter immédiatement les sources non fiables: ainsi si une source a un poids nul, elle ne perturbera pas les autres, puisqu'elle n'interviendra pas du tout dans le résultat.

La figure 2.13 montre le résultat pour l'exemple numérique. Il faut pour appliquer cette règle connaître les poids relatifs des différentes sources, dans cet exemple, les deux sources sont supposées aussi fiable l'une que l'autre. A cause de la contrainte 2.31, la seule possibilité est de prendre $\lambda_1 = \lambda_2 = 0.5$.

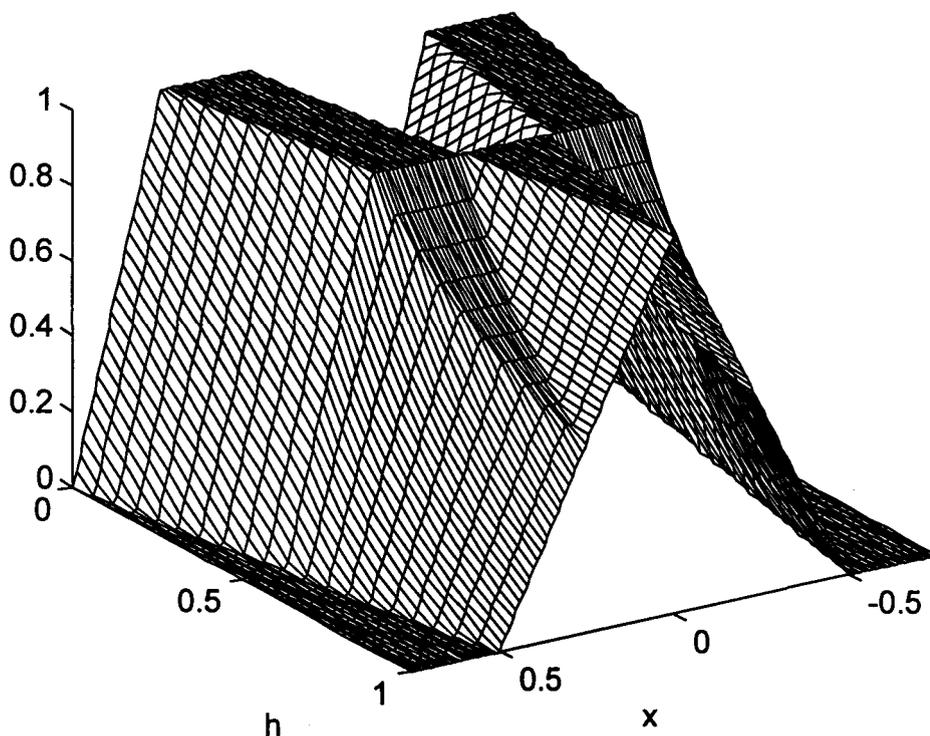


Fig. 2.13. Fusion linéaire.

Cette règle n'a pas de rapport avec les opérateurs logiques, comme le montre le résultat pour $h = 1$. Il est en effet totalement différent de l'habituelle intersection.

2.3.2. Fusion adaptative de type logique

Ce paragraphe montre divers autres opérateurs qui furent conçus dans le même objectif que l'opérateur adaptatif de Dubois et Prade.

2.3.2.1 Opérateur adaptatif saturé

La règle de Dubois et Prade était adaptative, mais du fait de l'utilisation de la normalisation du mode conjonctif, la règle était encore discontinue. Il est possible de résoudre cette discontinuité en diminuant l'influence de la normalisation.

La règle adaptative de Dubois et Prade avait la forme suivante:

$$\pi_{DP}(x) = \max\left(\frac{\pi_{\cap}^*(x)}{h(\pi_{\cap}^*(x))}, \min(1-h, \pi_{\cup}(x))\right) \quad (2.32)$$

En fait tout le problème vient de la normalisation par la hauteur h , qui n'est pas définie dans le cas où h est égal à 0. Pour éviter ce problème en 0, il suffit d'introduire un paramètre $s \in [0,1]$ appelé seuil ou saturation, de telle sorte qu'en dessous de cette valeur s , la normalisation soit

bloquée: le terme conjonctif $\frac{\pi_{\cap}^*(x)}{h(\pi_{\cap}^*(x))}$ est alors remplacé par:

$$\pi_{\cap \text{seuil}}^*(x) = \frac{\pi_{\cap}^*(x)}{\max(s, h(\pi_{\cap}^*(x)))} \quad (2.33)$$

Si s est strictement supérieur à 0, la discontinuité disparaît, si s est égal à 0 le mode conjonctif classique est retrouvé. Cependant le terme défini par 2.33 n'est pas normalisé si $h < s$, sa hauteur étant égale à $h/\max(h,s)$. Le déficit de normalisation vaut alors $1-h/\max(h,s)$.

S'il était normalisé par une relation du type:

$$\pi_{\cap \text{seuil}}(x) = \frac{\pi_{\cap \text{seuil}}^*(x)}{h(\pi_{\cap \text{seuil}}^*(x))} = \frac{\pi_{\cap \text{seuil}}^*(x) \max(h,s)}{h} \quad (2.34)$$

les problèmes de discontinuité de l'intersection seraient retrouvés dans le cas où $h=0$.

L'idée est alors de reporter le déficit de normalisation $1-h/\max(h,s)$ sur le terme disjonctif en augmentant la valeur de $1-h$ dans 2.32. Un calcul montre que le terme $\frac{1-\max(h,s)}{1-s}$ vaut 1 si

$h < s$ donc le terme $\min\left(\frac{1-\max(h,s)}{1-s}, \pi_{\cup}(x)\right)$ vaut toujours $\pi_{\cup}(x)$ si $h < s$ (et donc ce terme est normalisé puisque l'union est toujours normalisée par construction).

La règle adaptative saturée est alors définie par [Deveughele et Dubuisson, 1996]:

$$\pi_{\text{adaseuil}}(x) = \max\left(\frac{\pi_{\cap}^*(x)}{\max(s, h)}, \min\left(\frac{1-\max(h,s)}{1-s}, \pi_{\cup}(x)\right)\right) \quad (2.35)$$

Deux résultats obtenus avec cette règle de fusion sont donnés aux 2.14 et 2.15, et par là même ces figures soulignent l'inconvénient de cette règle, qui est la difficulté du choix de s .

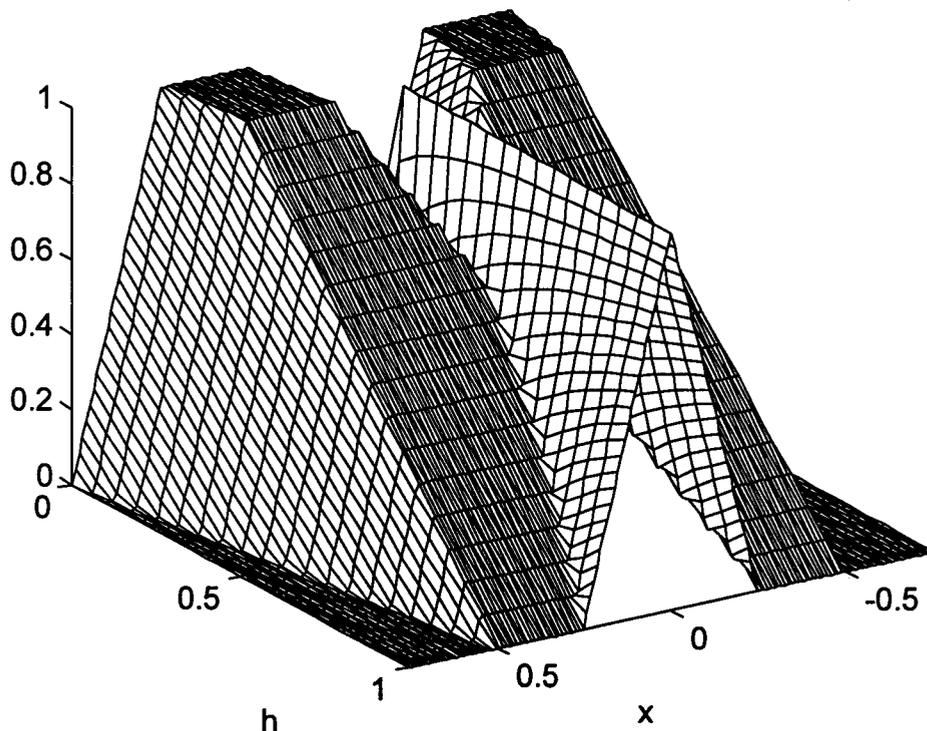


Fig. 2.14. Fusion adaptative saturée ($s=0.2$).

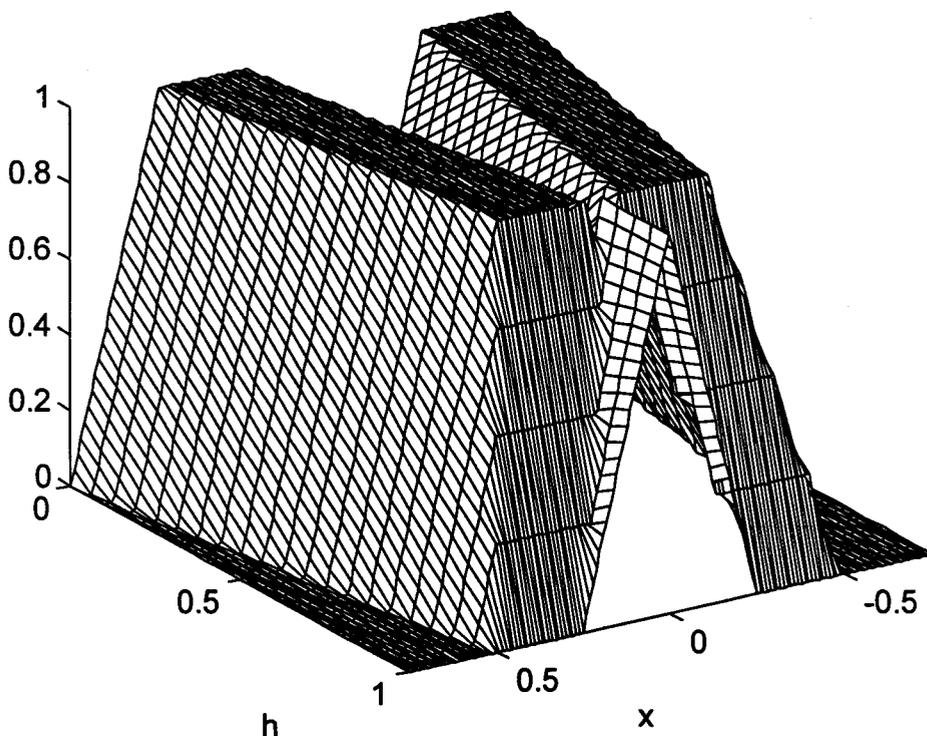


Fig. 2.15. Fusion adaptative saturée ($s=0.8$).

En effet s'il est vrai que la discontinuité a disparu et que cette règle permet un passage de l'intersection à l'union, le paramètre de seuil s joue un grand rôle, et il faut alors l'estimer précisément.

2.3.2.2 Opérateur adaptatif linéaire.

La fusion de Dubois et Prade repose sur une fusion logique entre d'une part une intersection normalisée et d'autre part une union, cette dernière étant saturée par une quantité dépendant du conflit entre les sources. Or il est possible de fusionner par une autre approche l'intersection et l'union: il suffit d'utiliser une combinaison linéaire des deux modes de fusion, avec des pondérations dépendant toujours de la contradiction entre les sources.

La construction est la suivante: la somme se compose de deux termes, $\pi_{\cap}(x)$ et $\pi_{\cup}(x)$. Le coefficient du premier terme est h : si $h = 1$, le conflit est nul, et il faut effectuer une intersection. Si h est égal à 0, le conflit est complet et il ne faut pas effectuer d'intersection.

La somme étant linéaire, le coefficient du second terme est donc $1-h$.

La règle est donc la suivante:

$$\pi_{\text{adalin}}(x) = \text{No}(h\pi_{\cap}(x) + (1-h)\pi_{\cup}(x))$$

Il peut y avoir une simplification sur le terme $h\pi_{\cap}(x)$, qui est égal à $\pi_{\cap}^*(x)$. La discontinuité disparaît donc d'elle même. La forme finale est donc

$$\pi_{\text{adalin}}(x) = \text{No}(\pi_{\cap}^*(x) + (1-h)\pi_{\cup}(x)) \quad (2.36)$$

La figure 2.17 montre le résultat pour l'exemple numérique.

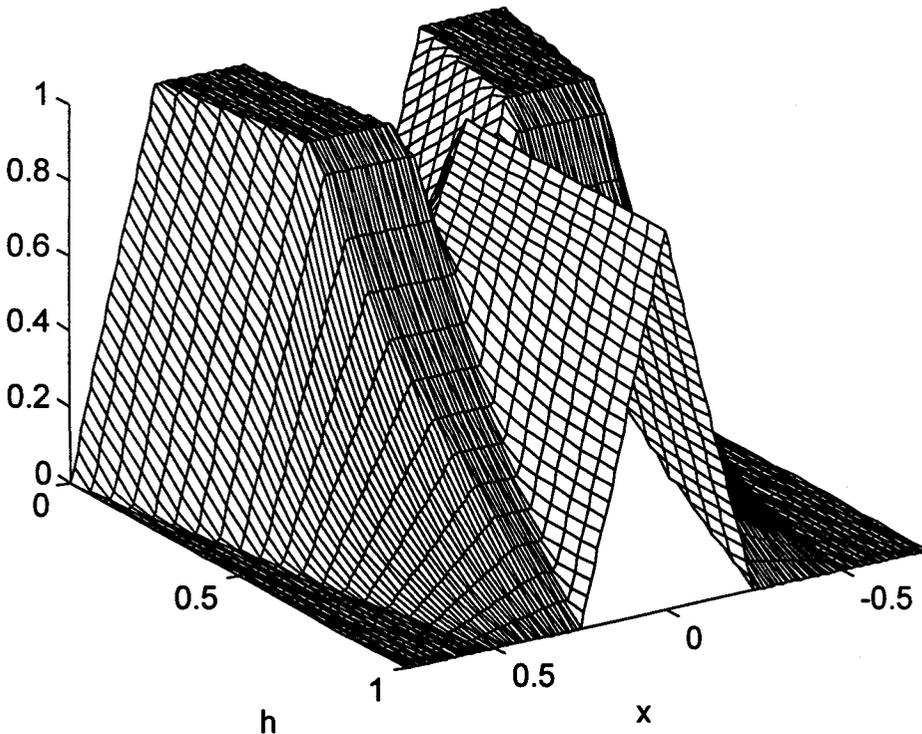


Fig. 2.17. Fusion adaptative linéaire.

Bien que cette règle ait été construite sur un principe de somme linéaire d'opérateurs, les variations de degré de possibilité ne sont pas linéaires. Elle privilégie l'intersection par rapport à l'union. En effet pour $h < 0.5$, le maximum de possibilité de la fonction est atteint pour $x=0$, et ensuite seulement le maximum de la fonction est atteint pour les x faisant partie des coeurs des trapèzes.

Dans le cas de l'exemple, cette règle ressemble à une fusion adaptative saturée avec un seuil s proche de 0.4.

Plutôt que de fusionner linéairement l'intersection normalisée $\pi_{\cap}(x)$ et l'union $\pi_{\cup}(x)$, il est possible de préférer fusionner linéairement l'intersection non normalisée $\pi_{\cap}^*(x)$ et l'union $\pi_{\cup}(x)$.

La relation est dans ce cas

$$\pi_{\text{adalin}}(x) = \text{No}(h\pi_{\cap}^*(x) + (1-h)\pi_{\cup}(x)) \quad (2.37)$$

Dans ce cas la transition entre l'intersection et l'union se produit exactement pour $h=0.5$. La figure 2.18 montre le résultat.

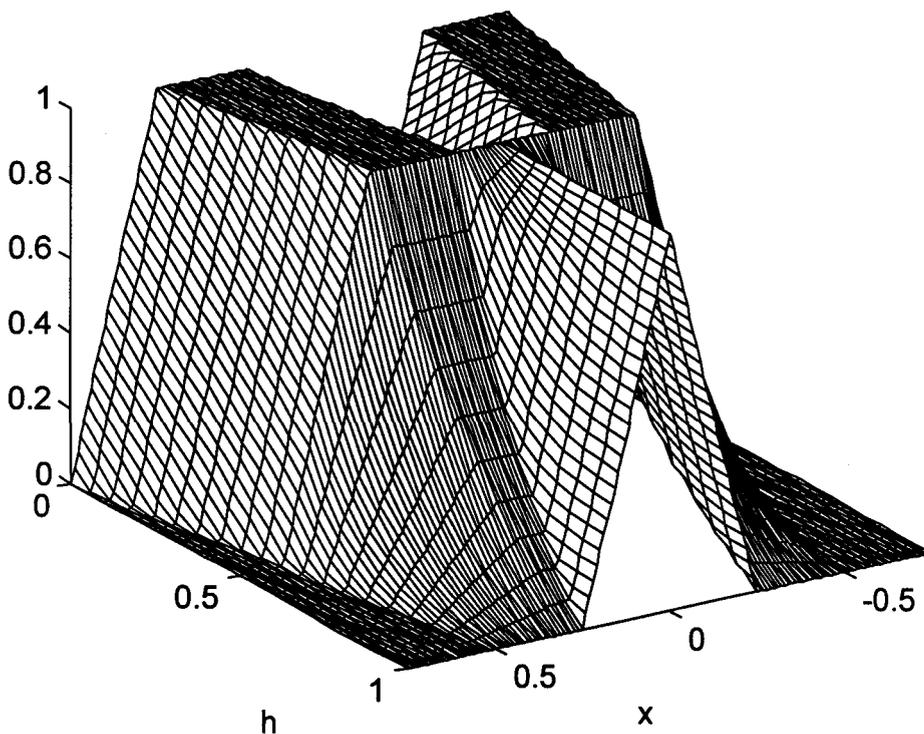


Fig. 2.18. Fusion adaptative linéaire avec transition forcée en $h=0.5$.

2.3.2.3 Opérateur de type adaptatif hybride

Ceci est un opérateur de type un peu différent des précédents, puisqu'il effectue principalement une fusion de type disjonctive. L'application de l'union suppose qu'une source fiable soit cachée parmi d'autres sources. Afin de conserver cette source, toutes les informations sont englobées, mais ce faisant l'information finale est très peu précise.

Mais il se pourrait que plusieurs sources fiables soient cachées parmi d'autres non fiables. Dans ce cas il serait intéressant d'en prendre l'intersection, puis de les unir aux autres. C'est de cette manière que fonctionne l'opérateur adaptatif hybride: parmi n sources, j sont supposées fiables. L'opérateur est alors défini par:

$$\pi_j(x) = \max_{\substack{J, \text{Card}(J)=j \\ J \subset \{1..n\}}} \min_{k \in J} \pi_k(x) \quad (2.38)$$

Le problème de cet opérateur est qu'il faille estimer le nombre j de sources fiables. Il est possible de définir un critère permettant d'estimer un nombre optimal de sources fiables. En effet la quantité $h_j^* = \max_{\substack{J, \text{Card}(J)=j \\ J \subset \{1..n\}}} \left(\max_x \min_{k \in J} \pi_k(x) \right)$ calcule la meilleure hauteur possible lorsque

l'intersection entre j sources est calculée. Cependant cette quantité a un optimum global, pour $j=1$, avec lequel h_1^* vaut 1, la fusion devenant alors une union de toutes les sources, une seule étant considérée fiable à chaque fois.

Il faut alors aussi chercher à maximiser le nombre de sources fiables, et donc il faut maximiser le nombre j dans le critère.

Le critère est donc une somme de deux composantes, qu'il faut rendre commensurables. Il est possible de choisir le critère suivant:

$$f(j) = \max_{\substack{J, \text{Card}(J)=j \\ J \subset \{1..n\}}} \left(\max_x \min_{k \in J} \pi_k(x) \right) + j/n \quad (2.39)$$

Le nombre optimal de sources est alors le j^* tel que

$$j^* = \text{Arg}(\max(f_j)) \quad (2.40)$$

2.3.3 Opérateurs de compromis

Ces opérateurs sont plus particulièrement utilisés en analyse de la décision, lorsque des informations de fiabilité moyenne sont fusionnées en grand nombre [Chen and Hwang, 1992]. Dans ce cas, plutôt que de construire une distribution globale sur des bases logiques, qui soit privilégient les points communs, soit englobent toutes les informations, des règles de fusion de type compromis sont utilisées.

Elles ont pour but de dégager l'opinion moyenne entre les sources d'informations, et font appel à des opérateurs de moyenne venant de l'analyse classique. Les opérateurs de compromis définis au premier chapitre peuvent être aussi utilisés.

Un opérateur de moyenne M est un opérateur qui respecte les propriétés suivantes:

$$M: [0,1]^n \rightarrow [0,1]$$

$$- M(x, x, \dots, x) = x$$

- M est commutatif

- M est non décroissant pour chaque argument.

Parfois une condition de continuité est appliquée. Certains chercheurs enlèvent d'office les opérateurs min et max de cette famille [Grabish, 1995b].

Des exemples en sont:

$$\begin{aligned}
 M(x) &= \frac{1}{n} \sum_{i=1}^n \pi_i(x) \\
 M(x) &= \left(\frac{1}{n} \sum_{i=1}^n \pi_i^{-1}(x) \right)^{-1} \\
 M(x) &= \left(\frac{1}{n} \sum_{i=1}^n \pi_i^\alpha(x) \right)^{1/\alpha} \\
 M(x) &= f^{-1} \left(\frac{1}{n} \sum_{i=1}^n f(\pi_i(x)) \right) \tag{2.41}
 \end{aligned}$$

avec f une fonction continue strictement monotone

Il est possible de pondérer chacune des distributions avec un poids. Dans l'exemple de la moyenne arithmétique, le résultat est celui étudié au paragraphe 2.3.1.4.

Parmi les opérateurs de compromis se trouvent aussi l'opérateur médian, qui consiste à trier les degrés de possibilités et à renvoyer comme résultat la valeur médiane de la liste:

Pour un nombre impair $2n-1$ de sources, le résultat est ainsi:

$$\text{Med}(x) = \pi_{(n)x}(x) \tag{2.42}$$

Avec pour chaque x la liste des sources triées de telle sorte que $\pi_{(1)x}(x) \leq \dots \leq \pi_{(2n-1)x}(x)$, l'indice $(2n-1)x$ indiquant une permutation des indices pour un x donné.

Cet opérateur peut demander beaucoup de temps de calcul, puisque pour un espace X de cardinal q il faut trier q fois des listes de $2n-1$ termes. Yager a généralisé cet opérateur médian pour définir un opérateur médian moyen: il est noté OWA pour ordered weighted average et permet de retrouver un grand nombre d'autres opérateurs:

$$\text{OWA}_{\lambda_i}(x) = \sum_{i=1}^{2n-1} \lambda_i \pi_{(i)x}(x) \tag{2.43}$$

avec $\sum \lambda_i = 1$.

Dans cet opérateur les coefficients λ_i ne sont pas associés à une source en particulier (ce ne sont pas des indices de fiabilité) mais à un rang en particulier, pour attacher plus d'importance au degré de possibilité le plus important, ou au contraire au degré de possibilité le plus faible, suivant les choix de l'utilisateur, plus ou moins pessimiste ou optimiste... Les opérateurs OWA alliés au principe d'extension appliqué à des règles de fusion forment un ensemble récent de solutions possibles à la fusion utilisant la fiabilité des sources [Kelman et Yager, 1997].

Ainsi pour $\lambda = [1, 0, \dots, 0]$, l'opérateur min est retrouvé, puisqu'en chaque point c'est le plus petit degré qui est conservé. Pour $\lambda = [0, \dots, 0, 1]$ l'opérateur max est retrouvé. Pour $\lambda_i = 1/(2n-1)$, la moyenne arithmétique est retrouvée.

La famille des opérateurs S-OWA (pour Soft OWA) est dérivée des opérateurs OWA lorsque l'opérateur OWA est égal à la moyenne arithmétique. Le Soft vient du fait que l'opérateur S-

OWA est une moyenne réalisée à partir de la moyenne arithmétique et d'un opérateur logique. Etant donné un coefficient $\alpha \in [0,1]$, un opérateur S-OWA $_{\alpha}$ est ainsi défini par:

$$S - OWA_{\alpha}(x) = \alpha \pi_{\alpha}^*(x) + (1 - \alpha) \sum_{i=1}^n \pi_i(x) / n \quad (2.44)$$

Cette relation est la forme non normalisée, car cet opérateur est aussi utilisé en logique floue pure, et notamment dans un contrôleur flou, pour lequel le paramètre α est un paramètre à optimiser [Yager et al, 1994].

La figure 2.19 donne le résultat de l'application de cet opérateur dans l'exemple de la fusion des distributions π_1 et π_2 .

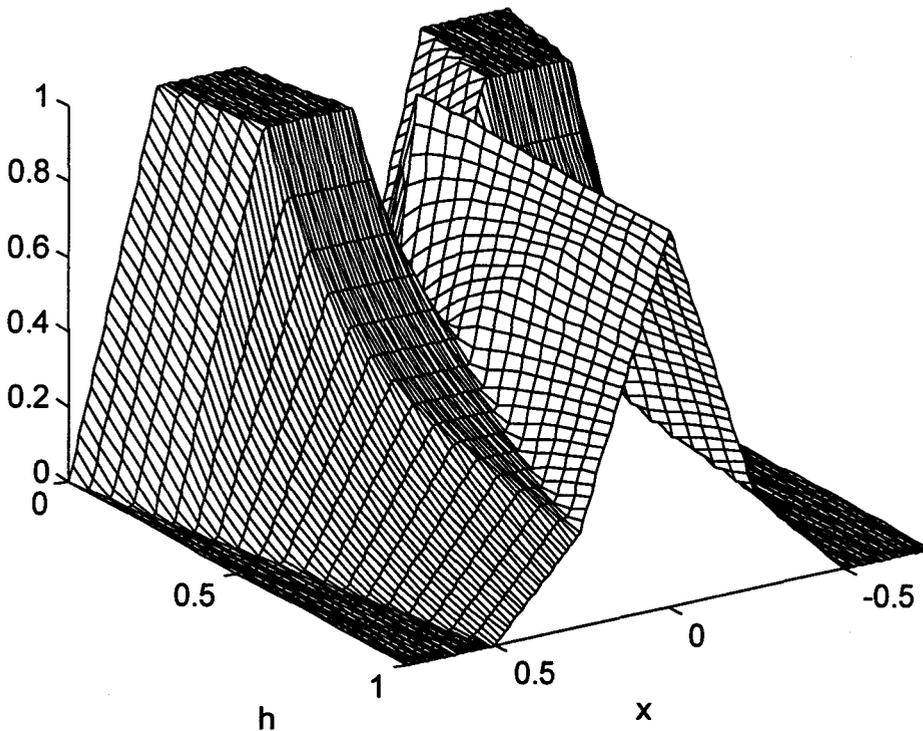


Fig. 2.19. Fusion basée sur opérateur S-OWA $_{\alpha}$ pour $\alpha=0.5$.

Cet opérateur est continu à cause du terme de la moyenne arithmétique. Il ne converge pas vers l'intersection lorsque les sources sont en accord ($h = 1$).

2.3.4 Intégrales floues

La théorie des probabilités est fondée sur la théorie de l'intégration, qui fait elle même appel à la théorie de la mesure.

La théorie de l'intégration floue a été définie pour généraliser à des mesures non additives la théorie de l'intégration des fonctions classiques [Grabish, 1995a]

Cette théorie fait appel à de solides développements mathématiques qui ne peuvent être détaillés dans ces pages. Dans une perspective de fusion de données, seules quelques définitions seront données dans le cadre d'intégrales discrètes.

Tout d'abord est définie une mesure floue $\mu: 2^X \rightarrow [0,1]$, avec 2^X l'ensemble des parties de X . μ est une mesure floue si et seulement si:

$$\mu(\emptyset) = 0, \mu(X) = 1, \text{ et si } A \subset B; \text{ alors } \mu(A) \leq \mu(B)$$

Les mesures de possibilités sont des exemples de mesures floues.

Pour comparer avec la théorie des probabilités, la propriété d'additivité $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$ a été abandonnée pour une propriété plus faible de monotonie.

$(X, 2^X, \mu)$ est dit un espace de mesure floue.

Il est alors possible de définir l'intégrale discrète de Sugeno de $f: X \rightarrow [0,1]$ par rapport à la mesure μ :

$$S_\mu(f(x_1), \dots, f(x_n)) = \bigcup_{i=1}^n f(x_{(i)}) \cap \mu(A_{(i)}) \quad (2.45)$$

\bigcup et \cap sont deux opérateurs d'union et d'intersection, par exemple max et min.

La notation $x_{(i)}$ indique une permutation sur les indices de telle sorte que $0 \leq f(x_{(1)}) \leq \dots \leq f(x_{(n)}) \leq 1$, et l'ensemble $A_{(i)}$ est défini par $A_{(i)} = \{x_{(i)}, \dots, x_{(n)}\}$.

Par exemple pour $n=3$,

$$S_\mu(0.6, 0.2, 0.8) = (0.2 \cap \mu(\{x_2, x_1, x_3\})) \cup (0.6 \cap \mu(\{x_1, x_3\})) \cup (0.8 \cap \mu(\{x_3\}))$$

avec \cap une t-norme (par exemple min) et \cup une t-conorme (par exemple max), et la notation $0.8 \cap \mu(\{x_3\}) \equiv \cap(0.8, \mu(\{x_3\}))$.

L'intégrale discrète de Choquet de $f: X \rightarrow [0,1]$ par rapport à la mesure μ est définie par:

$$C_\mu(f(x_1), \dots, f(x_n)) = \sum_{i=1}^n (f(x_{(i)}) - f(x_{(i-1)})) \mu(A_{(i)}) \quad (2.46)$$

avec $f(x_{(0)})$ pris par convention égal à 0.

Ces deux définitions d'intégrales floues sont des cas particuliers d'une intégrale plus générale appelée intégrale floue de t-conorme d'une fonction f .

Quelques propriétés des intégrales floues sont mentionnées:

- elles sont monotones par rapport à la mesure (si $\mu \leq \mu'$ alors $C_\mu \leq C_{\mu'}$ par exemple).
- elles sont monotones par rapport à la fonction à intégrer (si $f \leq f'$ alors $C_\mu(f) \leq C_\mu(f')$ par exemple).
- elles sont comprises entre l'opérateur max et opérateur min.
- si la mesure μ est additive, l'intégrale de Lebesgue ou de Stieljés est retrouvée:

$$C_{\mu_{\text{addi}}}(f(x_1), \dots, f(x_n)) = \sum_{i=1}^n f(x_i) \mu_{\text{addi}}(x_i)$$

Les intégrales floues apparaissent comme des opérateurs qui englobent un grand nombre d'autres, en particulier les opérateurs OWA [Grabish, 1995b].

Cependant elles ne permettent pas de retrouver les t-conormes et t-normes (sauf le max et le min), et ne peuvent pas non plus approximer n'importe quel opérateur flou au sens usuel.

2.3.5 Fusion basée sur les degrés de confiances.

Tous les opérateurs présentés jusqu'à présents, à part l'opérateur de fusion linéaire, ne privilégiaient pas une source par rapport à une autre. Elles étaient toutes considérées de la même façon, seul éventuellement le mode du fusion changeait.

L'opérateur de fusion linéaire est le seul à présenter la caractéristique de pouvoir renforcer une source par rapport à une autre, par l'introduction de coefficients de fiabilité ou de confiance.

Dans ce paragraphe sont étudiés d'autres opérateurs présentant cette capacité. Indépendamment du mode de fusion considéré, tous permettent de prendre en compte des informations supplémentaires, qui ne sont plus du domaine de l'évaluation des possibilités des événements, mais qui concernent en les quantifiant les qualités des sources. L'inconvénient des ces approches est de nécessiter la connaissance de ces coefficients de confiance.

2.3.5.1 Opérateur de priorité

Cet opérateur est un peu particulier en cela qu'il ne nécessite pas de connaissance précise sur les fiabilités des sources, mais simplement un ordre sur ces fiabilités. Par exemple la source 1 est plus fiable que la source 2.

Dans ce cas le principe est de ne prendre en compte la deuxième source que si elle n'est pas en contradiction avec la première. Cette contradiction est utilisée par l'intermédiaire de la classique hauteur de l'intersection de 2 sources. La règle est:

$$\pi_{1 \rightarrow 2}(x) = \min(\pi_1(x), \max(\pi_2(x), 1 - h)) \quad (2.47)$$

Si $h = 0$, le résultat est π_1 , si $h = 1$, le résultat est $\pi_1 \cap \pi_2$. La figure 2.20 montre le résultat.

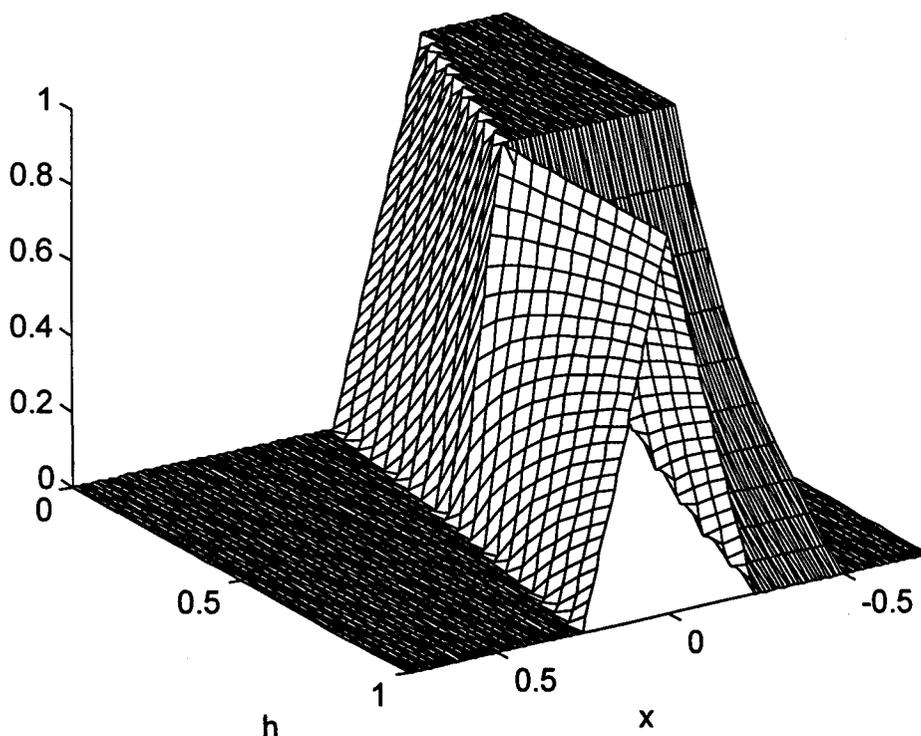


Fig. 2.20. Fusion avec priorité de π_1 sur π_2 .

2.3.5.2 Fusion linéaire avec coefficients de confiance

Cette règle de fusion a déjà été examinée au paragraphe 2.3.1.4. C'est le premier exemple de règle de fusion où des degrés quantifiant la qualité d'une source de fusion apparaissent.

Dans cette règle, cependant, ces coefficients sont liés par la contrainte $\sum \lambda_i = 1$. Aussi ces coefficients ne sont définis que les uns par rapport aux autres, et ne peuvent être définis de manière absolue. Chaque λ_i ne peut être défini indépendamment avec la seule contrainte que $\lambda_i \in [0, 1]$. Ainsi un coefficient de confiance obtenu pour un problème donné ne peut pas être utilisé dans un autre problème où la source en question serait fusionnée avec d'autres sources que les premières. Cette méthode de fusion apparaît néanmoins très riche et très souple.

2.3.5.3 Fusion basée sur les transformations d'incertitude

Dans cette approche, contrairement à la précédente, les degrés de fiabilité sont utilisés sans contrainte. Ce sont simplement des coefficients compris entre 0 et 1: $\lambda = [\lambda_1 \dots \lambda_n] \in [0, 1]^n$.

L'idée est de tenir compte de ces coefficients, non pas dans une étape de fusion, mais dans une étape de transformation des sources en fonction de leur fiabilité, afin d'obtenir des sources fiables. A cet effet, les transformations 2.13 et 2.14 données au paragraphe 2.2.6.3 peuvent être utilisées. Les sources transformées sont alors supposées parfaitement fiables, et donc fusionnées par une règle conjonctive.

Par exemple avec une transformation de Dubois et Prade, les résultats sont:

$$\pi_{\text{TransfDP}}(x) = \text{No}(\min(\max(\pi_1(x), 1 - \lambda_1), \dots, \max(\pi_n(x), 1 - \lambda_n))) \quad (2.48)$$

La figure 2.21 montre le résultat en appliquant la transformation de Dubois et Prade. Les deux coefficients de fiabilité sont pris égaux à h.

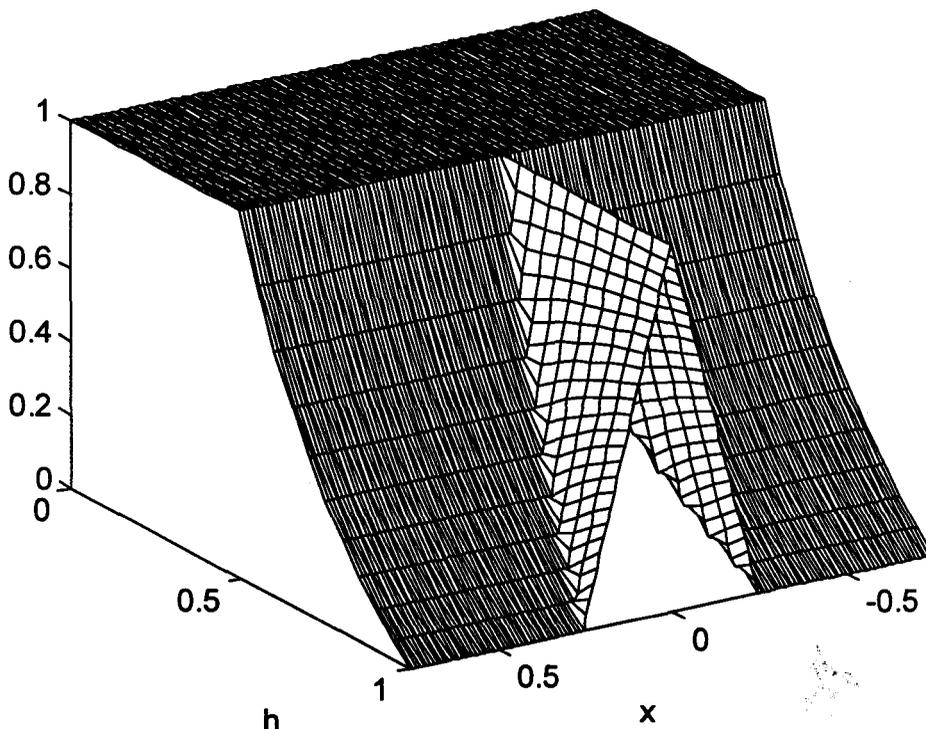


Fig. 2.21. Fusion avec degrés de confiance. $\pi_{\text{TransfDP}}(x) = [T_{\text{DP}}(\pi_1(x), h)] \cap [T_{\text{DP}}(\pi_2(x), h)]$.

La figure 2.22 montre la même fusion en appliquant la transformation de Yager (2.14).

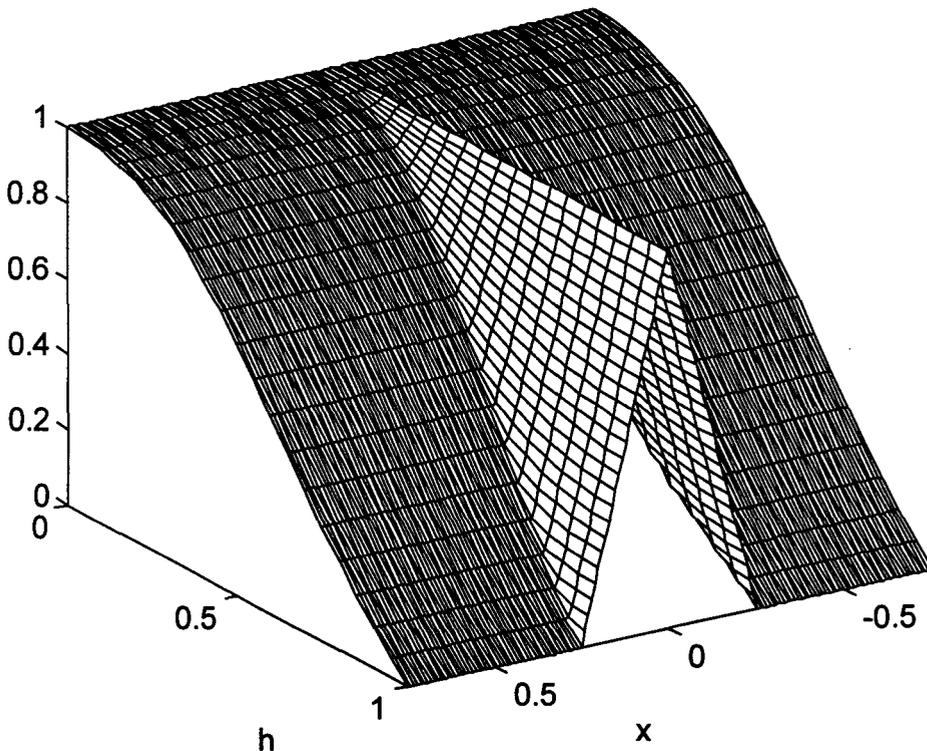


Fig. 2.22. Fusion avec degrés de confiance. $\pi_{\text{Fiabilité Yag}}(x)=[T_{\text{Yag}}(\pi_1(x),h)] \cap [T_{\text{Yag}}(\pi_2(x),h)]$.

L'inconvénient de ces deux approches est que le support du résultat devient immédiatement l'ensemble du référentiel X dès qu'une seule source n'est plus absolument fiable (autrement dit dès que le vecteur des fiabilités des sources est différent de $[1...1]$).

Pour la plupart des autres modes de fusion, le résultat était toujours compris entre l'intersection et l'union, et donc pour la plupart des autres fusions, le support du résultat était toujours inclus dans le support de l'union des sources.

La transformation de Dubois et Prade semble donner les résultats les moins précis des deux approches, puisque le résultat est la complète ignorance dès que h est inférieur à 0.5. Ceci est dû à la forme particulière des deux trapèzes, mais il est vrai que la transformation 2.13 nivelle plus les informations que la transformation 2.14: pour des degrés de possibilités $\pi(x)$ inférieurs à $1-h$, la transformation 2.13 renvoie toujours $1-h$ comme valeur finale, tandis que la transformation 2.14 conserve plus d'information, et pour deux degrés $\pi(x)$ et $\pi(x')$ différents mais inférieurs à $1-h$, les degrés après transformation sont différents:

$$T_{\text{Yag}}(\pi(x),h) \neq T_{\text{Yag}}(\pi(x'),h) \text{ si } \pi(x) \neq \pi(x')$$

$$T_{\text{DP}}(\pi(x),h) \neq T_{\text{DP}}(\pi(x'),h) \text{ si } \pi(x) \neq \pi(x') \text{ et si } (\pi(x) \geq 1-h \text{ ou } \pi(x') \geq 1-h).$$

2.3.6 Conclusion sur les modes de fusion

Un grand nombre d'opérateurs de fusion ont été envisagés. Mais seuls quatre grandes familles sont à retenir.

Il y a en premier lieu les opérateurs logiques, tels que l'union et l'intersection, qui ont des comportements naturels suivants les hypothèses de fiabilité des sources. Dans cette famille il

Il y a des opérateurs statiques (l'union et l'intersection), et des opérateurs dynamiques ou adaptatifs, qui évoluent entre l'intersection et l'union suivant un indice global de fiabilité des sources.

Il y a en second lieu des opérateurs de compromis, qui ne sont pas forcément liés aux opérateurs logiques, et qui donnent des résultats moyens lorsque des informations moyennement fiables sont agrégées. La moyenne arithmétique en est un exemple.

Il y a en troisième lieu les intégrales floues, qui peuvent être utilisées comme opérateurs de fusion. En fait ces intégrales permettent d'englober une partie des autres familles de fusion et ont des comportements divers suivant la définition exacte de l'intégrale.

Enfin il y a la famille des opérateurs basés sur un degré de fiabilité des sources. Ceux-ci se distinguent des autres opérateurs parce qu'ils permettent de prendre en compte de manière quantitative une connaissance sur les degrés de fiabilités des sources. Cependant ils n'ont pas en général de comportement logique. La fusion linéaire en est un exemple.

2.4 Un nouvel opérateur de fusion

2.4.1 Objectifs du nouvel opérateur

Aucun des opérateurs étudiés au paragraphe 2.3 n'apparaît universel. Tous ont des conditions d'applications spécifiques, et les opérateurs adaptatifs ne permettent pas d'accorder plus de poids à telle ou telle source de données. Les opérateurs basés sur les degrés de confiance ont cette capacité mais n'ont pas de comportements logiques.

Aussi s'est fait sentir le besoin d'un nouvel opérateur qui prendrait en compte la fiabilité des sources tout en ayant un comportement logique, ainsi suivant les degrés respectifs de confiance, le mode de fusion serait soit conjonctif, soit disjonctif ou encore de compromis, tout en permettant de rejeter les sources non fiables. De plus il serait souhaitable de ne pas introduire de contrainte sur les degrés de confiances.

Avec les précédents modes de fusion, il n'y avait aucune estimation de la qualité du résultat obtenu, et par exemple avec les méthodes basées sur la transformation des sources selon leur degré de qualité, l'intersection était considérée parfaite et le résultat était considéré parfaitement fiable, même si les sources à fusionner étaient peu fiables.

Il serait intéressant de pouvoir quantifier, au même titre que les degrés de fiabilité des sources, la fiabilité du résultat obtenu.

Avec les méthodes de fusion basées sur la fiabilité se pose le problème de l'estimation des degrés de confiance accordée aux sources. Ce sujet n'a été que rarement abordé dans la littérature. Ce paragraphe a donc pour objectif de présenter un ensemble d'outils de fusion répondant à ces trois problèmes.

Dans une première partie sera exposée la construction du nouvel opérateur, construction qui en fait conduit à trois nouvelles versions. Une méthode d'estimation des coefficients de confiance sur lesquels est basé le nouvel opérateur sera exposée dans une seconde partie. Enfin sera donnée une tentative d'estimation du degré de confiance qu'il est possible d'accorder au résultat calculé.

2.4.2 Construction du nouvel opérateur.

2.4.2.1 Structure générale

Il faut d'abord présenter les informations à fusionner: N distributions de possibilités π_i , chacune associée à un degré de confiance $t_i \in [0,1]$. Une source de degré de confiance t_i égal à 0 est considérée comme totalement non fiable et une source de degré de confiance t_i égal à 1 comme totalement certaine. Le vecteur de fiabilité t est constitué de l'ensemble des degrés de fiabilité des sources.

Comme dans l'approche de fusion linéaire, chaque source aura une importance croissant linéairement en fonction de son degré de confiance. Les informations élémentaires à fusionner seront donc des $\pi_i' = t_i \pi_i$ [Delmotte et al, 1995a].

Ensuite comme pour les règles adaptatives, il y aura une combinaison d'un mode conjonctif et d'un mode disjonctif. L'approche linéaire a été retenue.

Il y aura donc des coefficients estimant dans quelle mesure il faut appliquer le mode conjonctif par rapport au mode disjonctif, et ces coefficients dépendront de l'ensemble des coefficients de confiance.

La forme la plus simple possible est celle d'un polynôme du premier ordre dont les arguments sont les opérateurs conjonctifs et disjonctifs.

La structure générale est donc de la forme:

$$\pi(x) = \alpha(t_i) \max_i(t_i \pi_i(x)) + \beta(t_i) \min_i(t_i \pi_i(x)) + \chi(t_i) \max_i(t_i \pi_i(x)) \min_i(t_i \pi_i(x)) + \delta(t_i) \quad (2.50)$$

Le terme $\max_i(t_i \pi_i(x)) \min_i(t_i \pi_i(x))$ n'ayant pas un grand sens physique, le coefficient $\chi(t_i)$ est pris égal à 0.

Il reste donc à déterminer les coefficients α , β et δ .

Pour cela des cas limites seront utilisés, comme par exemple le comportement souhaité lorsque toutes les sources sont fiables ($t = [1 \dots 1]$).

2.4.2.2 Conditions limites

Les conditions limites sont obtenues dans le cas de coefficients de crédibilité binaire ($t_i \in \{0,1\}$).

Dans ce cas il est possible de classer les sources d'informations en deux classes: celles absolument fiables F et celles absolument non fiables NF .

Les conditions limites sont alors les suivantes:

- $NF = \emptyset$. Dans ce cas toutes les sources sont fiables. Une intersection est donc souhaitée.
- $NF \neq \emptyset$ et $F \neq \emptyset$. Dans ce cas il existe à la fois des sources fiables et des sources non fiables. Une union s'avère donc comme le mode de fusion le plus approprié
- $F = \emptyset$. Dans ce cas aucune source n'est fiable. Contrairement aux précédents cas il n'existe pas de comportement meilleur que les autres. Il est possible de préférer que le résultat

tende vers l'ignorance complète. Il est possible de préférer aussi que le résultat soit une simple union. Enfin il est possible à ce stade de ne pas implanter de comportement particulier, et souhaiter que le résultat avant normalisation tende vers $\pi(x)=0 \forall x$ (le complément de l'ignorance complète).

En fait les trois comportements envisagés dans le dernier cas conduisent à trois versions différentes du nouvel opérateur.

2.4.2.3 Version π_{Xor} du nouvel opérateur

La version développée correspond au cas limite où rien n'est mathématiquement imposé dans le cas où aucune source n'est fiable. L'indice Xor caractérise un coefficient particulier.

Les trois tables de vérité suivantes permettent de calculer les valeurs des trois coefficients α , β et δ représentant les coefficients de l'union, de l'intersection et de l'incertitude.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Union s'applique	Non	Oui	Non

Table 2.3. Cas limites utilisés pour déterminer le coefficient α de l'union.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Intersection s'applique	Non	Non	Oui

Table 2.4. Cas limites utilisés pour déterminer le coefficient β de l'intersection.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Incertain s'applique	Non	Non	Non

Table 2.5. Cas limites utilisés pour déterminer le coefficient δ de l'incertitude.

Ces trois tables permettent de calculer les coefficients α , β et δ comme des opérateurs logiques des ensembles F et NF, et donc des coefficients t_i binaires.

Il apparaît donc que le coefficient α est un Xor logique, que le coefficient β est un And, et que le coefficient δ est nul.

Cette approche est généralisée pour des coefficients de crédibilité t_i non binaires par l'intermédiaire d'opérateurs flous.

Le And flou est donc une t-norme. Le Xor flou par contre est peu rencontré en logique floue. Il est construit à partir de la relation en logique classique: $Xor(.) = (Or(.))And(Not (And (.)))$, généralisée grâce à une Négation et à un And flous (le Or s'en déduit).

En logique floue, il existe, comme cela a été montré au premier chapitre, plusieurs versions possibles pour la négation et le And. Si les opérateurs probabilistes sont retenus, les coefficients valent par exemple:

$$\pi_{\text{Xor}}(x) = \text{No}(\text{Xor}(t_i) \max_i(t_i, \pi_i(x)) + \text{And}(t_i) \min_i(t_i, \pi_i(x))) \quad (2.51)$$

Pour renforcer l'intersection, il est possible d'introduire des opérateurs de compromis et de modifier la pondération des π_i dans le "min" en utilisant les transformations du 2.2.6.3:

$$\pi(x) = \text{No}\left[\left(1 - \frac{1}{n} \sum t_i\right) \max_i(t_i, \pi_i(x)) + \left(\frac{1}{n} \sum t_i\right)^2 \min\{\min_i(1 - t_i + t_i \pi_i(x)), \cup_i \text{Supp}(\pi_i)\}\right]$$

2.4.2.4 Version π_{Nand} du nouvel opérateur

La version développée correspond au cas limite où l'union est choisie dans le cas de sources totalement non crédibles.

Les trois tables de vérité suivantes permettent de calculer les valeurs des trois coefficients α , β et δ représentant les coefficients de l'union, de l'intersection et de l'incertitude.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Union s'applique	Oui	Oui	Non

Table 2.7. Cas limites utilisés pour déterminer le coefficient α de l'union.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Intersection s'applique	Non	Non	Oui

Table 2.8. Cas limites utilisés pour déterminer le coefficient β de l'intersection.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Incertain s'applique	Non	Non	Non

Table 2.9. Cas limites utilisés pour déterminer le coefficient δ de l'incertitude.

Il apparaît donc que le coefficient α est un Nand logique, que le coefficient β est un And, et que le coefficient δ est nul.

Comme dans le paragraphe précédent, les opérateurs flous sont utilisés, et un Nand se calcule comme étant 1-And.

Si les opérateurs probabilistes sont retenus pour la négation et le And, les coefficients valent par exemple:

$$\alpha = \text{Nand}(t_i) = 1 - \prod t_i \text{ et } \beta = \text{And}(t_i) = \prod t_i$$

Le résultat pour la version π_{Nand} est dans tous les cas:

$$\pi_{\text{Nand}}(x) = \text{No}(\text{Nand}(t_i) \max_i(t_i, \pi_i(x)) + \text{And}(t_i) \min_i(t_i, \pi_i(x))) \quad (2.52)$$

2.4.2.5 Version π_{Nor} du nouvel opérateur

La version développée correspond au cas limite où l'ignorance complète est choisie si toutes les sources sont non fiables.

Les trois tables de vérité suivantes permettent de calculer les valeurs des trois coefficients α , β et δ représentant les coefficients de l'union, de l'intersection et de l'incertitude.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Union s'applique	Non	Oui	Non

Table 2.10. Cas limites utilisés pour déterminer le coefficient α de l'union.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Intersection s'applique	Non	Non	Oui

Table 2.11. Cas limites utilisés pour déterminer le coefficient β de l'intersection.

Sources fiables	Aucune	En partie	Toutes
Sources non fiables	Toutes	En partie	Aucune
Incertitude s'applique	Oui	Non	Non

Table 2.12. Cas limites utilisés pour déterminer le coefficient δ de l'incertitude.

Il apparaît donc que le coefficient α est un Xor logique, que le coefficient β est un And, et que le coefficient δ est Nor logique.

Comme dans les deux précédents paragraphes, les opérateurs sont créés à partir d'opérateurs flous. Le Nor est calculé à partir de la relation booléenne: $\text{Nor}(\cdot) = (\text{And}(\text{Not}(\cdot)))$.

Avec les conventions précédentes, les coefficients valent par exemple:

$$\alpha = \text{Xor}(t_i) = (1 - \prod(1 - t_i))(1 - \prod t_i) \text{ et } \beta = \text{And}(t_i) = \prod t_i, \delta = \prod(1 - t_i)$$

Le résultat pour la version π_{Nor} est dans tous les cas:

$$\pi_{\text{Nor}}(x) = \text{No}(\text{Xor}(t_i) \max_i(t_i \pi_i(x)) + \text{And}(t_i) \min_i(t_i \pi_i(x)) + \text{Nor}(t_i)) \quad (2.53)$$

2.4.2.6 Comparaisons des opérateurs

Le comportement des ces opérateurs est étudié dans ce paragraphe, en prenant l'exemple de la fusion des deux trapèzes. Ces opérateurs nécessitent la connaissance des coefficients t_i de confiance. Comme dans le paragraphe concernant la fusion par degrés de confiance avec les transformations de Dubois-Prade et de Yager, ces coefficients sont pris égaux à h . Il est à remarquer que ce choix pour les coefficients de confiance n'est pas dû au hasard, mais est au contraire issu du paragraphe 2.4.4 concernant l'estimation de ces coefficients de confiance.

Les figures 2.23 et 2.24 montrent les résultats pour la version π_{Xor} pour le choix de deux familles de t-normes différentes pour les coefficients α et β .

La figure 2.23 correspond à la famille probabiliste et la figure 2.24 à la famille de Zadeh.

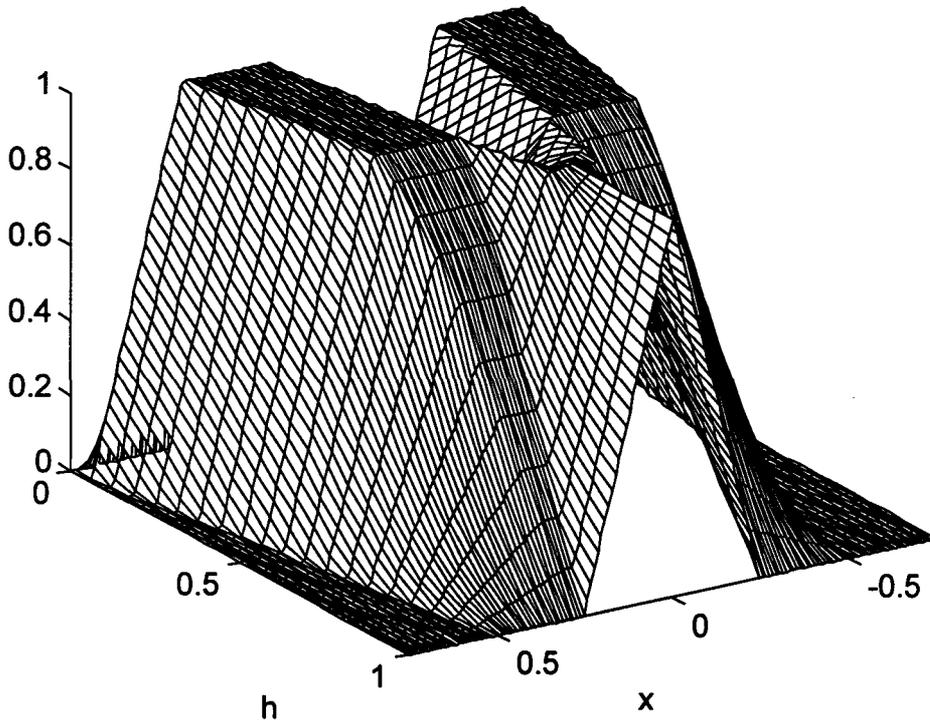


Fig. 2.23. Fusion avec la version π_{xor} du nouvel opérateur. Coefficients basés sur la famille probabiliste.

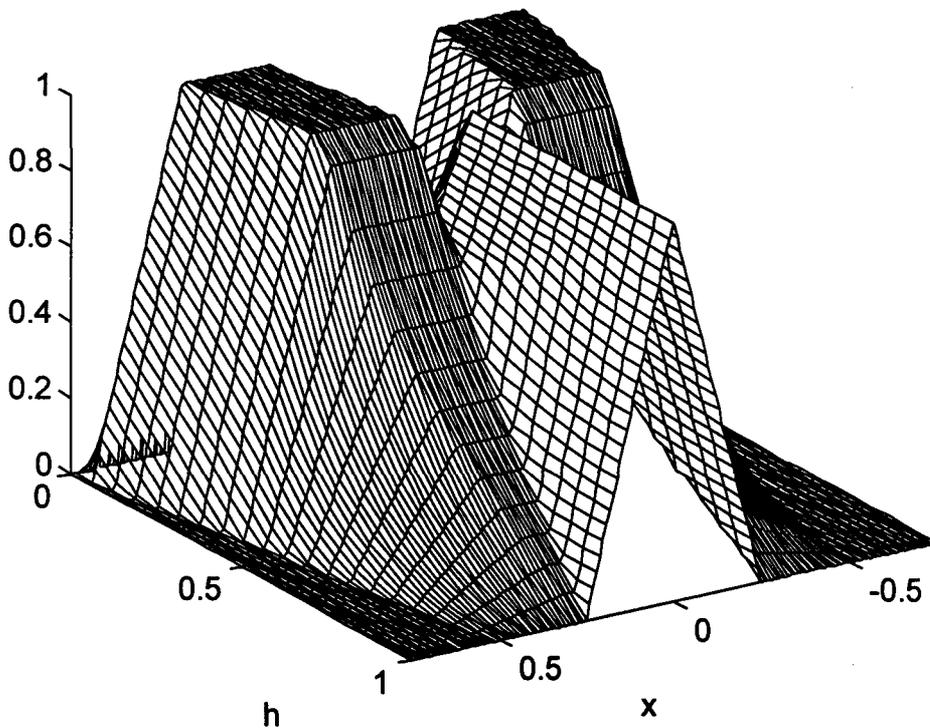


Fig. 2.24. Fusion avec la version π_{xor} du nouvel opérateur. Coefficients basés sur la famille de Zadeh.

Dans les deux cas l'opérateur tend vers l'union lorsque les sources sont toutes peu fiables. De plus il y a une continuité parfaite du résultat lorsque le conflit entre les sources devient absolu (lorsque $h=0$), mais ceci ne se produit dans le cas général que si les fiabilités des sources sont calculées comme dans le paragraphe sur l'estimation des degrés de crédibilité. Quelques explications sont données dans la comparaison avec la version π_{Nand} .

Il y a peu de différences entre les deux familles. En fait suivant les t-normes choisies, les comportements diffèrent sur la transition entre le comportement conjonctif et le comportement disjonctif. Dans le cas de la famille de Zadeh, le comportement disjonctif apparaît "plus tard" qu'avec la famille probabiliste. Cette dernière apparaît être un compromis parmi toutes les familles de t-normes (ainsi avec la t-norme discontinue, la plus petite des t-normes, le comportement est toujours disjonctif, sauf dans le cas où $h=1$ pour lequel l'intersection est choisie). En outre avec la famille probabiliste, la fusion paraît être plus continue qu'avec la famille de Zadeh. En effet avec la famille de Zadeh, lors de la transition mode conjonctif-mode disjonctif, le maximum de possibilité, atteint pour $x=0$ en mode conjonctif, passe brusquement aux valeurs de x qui appartiennent aux coeurs des trapèzes. Au contraire pour la famille probabiliste, le maximum de possibilité évolue continûment entre $x=0$ et les coeurs des deux trapèzes.

Pour la suite, à cause de ce comportement de compromis, la famille probabiliste est toujours choisie pour les coefficients α , β et δ .

La figure 2.25 montre le résultat obtenu avec la version π_{Nand} .

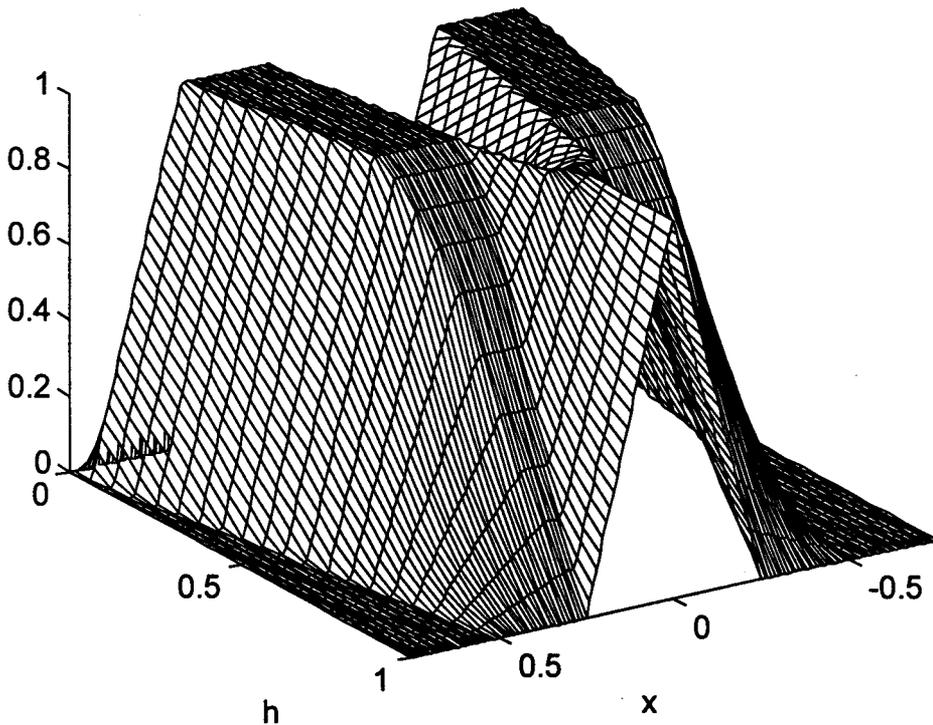


Fig. 2.25. Fusion avec la version π_{Nand} du nouvel opérateur. Coefficients basés sur la famille probabiliste.

En fait il y a peu de différences entre ces versions du nouvel opérateur, et ce qui est surprenant il n'y a aucune différence visible dans le cas où toutes les sources sont non fiables. Pourtant si l'opérateur π_{Nand} devait tendre vers l'union, cela n'avait pas été prévu pour la version π_{Xor} .

Pour comprendre les résultats, il faut effectuer une analyse fine au voisinage de $t_0=[0\dots 0]$. Pour simplifier, le vecteur t est pris égal à $[\varepsilon\dots\varepsilon]$, avec $\varepsilon \ll 1$.

Dans ce cas, le $\text{And}(t_i)$ est égal à ε^n .

Un développement limité du $\text{Nand}(t_i)$ au voisinage de t_0 est $1+o(\varepsilon)$

De même un développement limité du $\text{Xor}(t_i)$ au voisinage de t_0 est $n\varepsilon+o(\varepsilon^2)$.

Aussi dans les deux cas, au voisinage de t_0 le coefficient du terme disjonctif est un infiniment grand devant le coefficient du terme conjonctif, ce qui explique que les deux versions tendent vers l'union lorsque le vecteur des fiabilités tend vers le vecteur nul.

En fait des différences apparaissent plus nettement pour des valeurs de fiabilité intermédiaires, lorsque les fiabilités sont proches de 0.5. En effet le coefficient du terme disjonctif est beaucoup plus important pour la version π_{Nand} que pour la version π_{Xor} . La figure 2.26 montre la différence entre opérateur π_{Xor} et opérateur π_{Nand} .

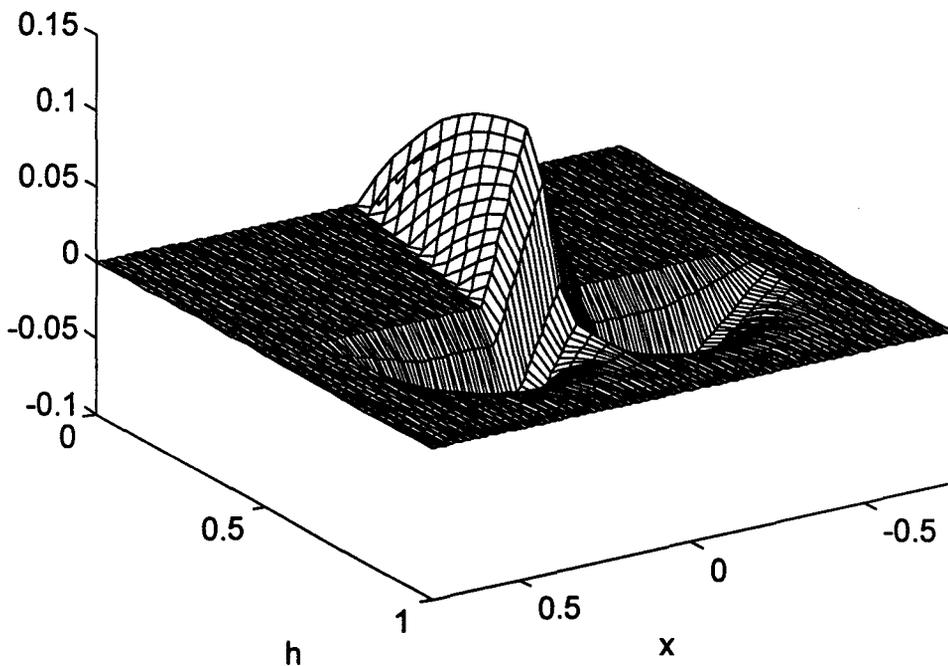


Fig. 2.26. Différence entre π_{Xor} et π_{Nand} .

Sur cette figure, la bosse correspond à des degrés de possibilités plus élevés pour le mode conjonctif (au voisinage de $x=0$), et les creux correspondent à des degrés de possibilités plus faibles pour le mode disjonctif (pour les x appartenant aux coeurs des deux trapèzes).

La version π_{Xor} privilégie donc dans tous les cas le mode de fusion conjonctif par rapport au mode disjonctif, et fournit donc des informations plus précises.

Cette version est donc préférée à la version π_{Nand} .

La figure 2.27 montre le résultat obtenu par l'application de la version π_{Nor} . Cette version, contrairement aux deux précédentes, ne tend pas vers l'union dans le cas de sources non crédibles.

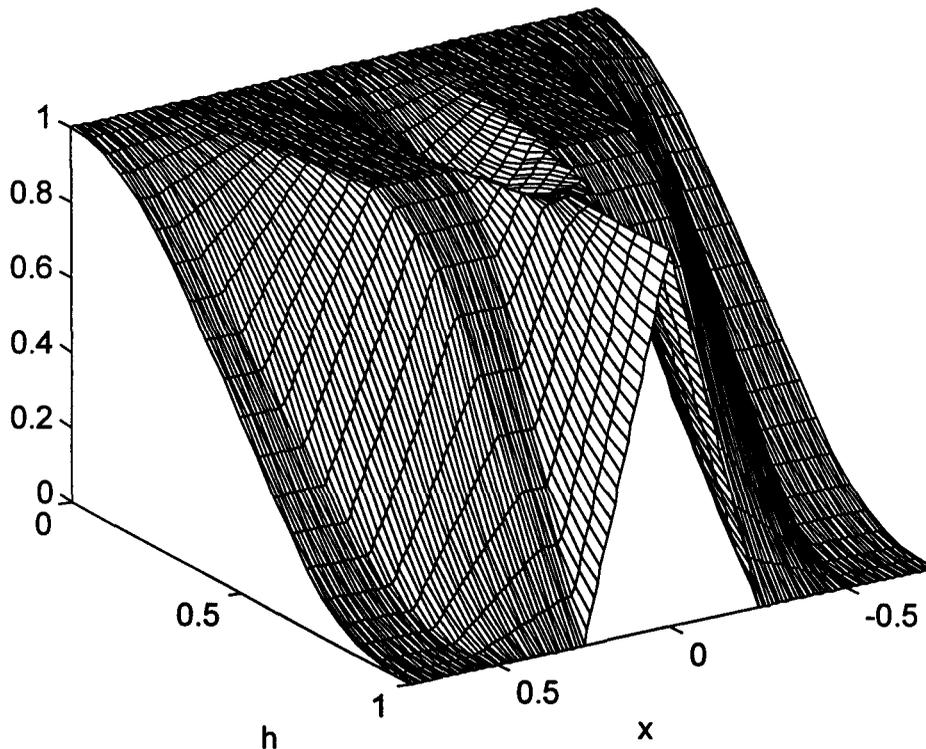


Fig. 2.27. Fusion avec la version π_{Nor} du nouvel opérateur. Coefficients basés sur la famille probabiliste.

Cet opérateur tend vers l'ignorance absolue lorsque les sources sont toutes non fiables. Dans l'exemple proposé, l'opérateur classique qui s'en rapproche le plus est celui de la fusion basée sur des degrés de confiance avec la transformation de Yager. Contrairement à ce dernier, pour des valeurs faibles de confiances ($t_i \leq 0.5$), les degrés de possibilités maximums sont atteints pour les x faisant partie des coeurs des deux trapèzes.

En conclusion ces versions du nouvel opérateur ont parfaitement rempli les objectifs assignés: avoir un comportement qui va de l'union à l'intersection (sauf la version π_{Nor} qui va de l'ignorance absolue à l'intersection), à partir de coefficients de confiance, sur lesquels pas ou peu de contraintes sont imposées a priori.

Dans l'exemple donné de la fusion de deux trapèzes dont le conflit varie, ces trois opérateurs ont eu des comportements analogues à des opérateurs déjà définis, comme par exemple l'opérateur adaptatif de Dubois et Prade pour π_{Xor} (avec cependant l'avantage de la continuité).

Ceci pourrait être trompeur, car le vecteur des confiances autorise en fait un nombre de comportements différents beaucoup plus large. Ainsi la fusion des deux trapèzes avec des coefficients constants pris égaux à 1 donne exactement l'intersection classique, comme le montre la figure 2.28. Dans ce cas les coefficients de confiance ne sont pas calculés comme dans le paragraphe 2.4.4 les estimant, et le résultat n'est pas continu (en $h=0$).

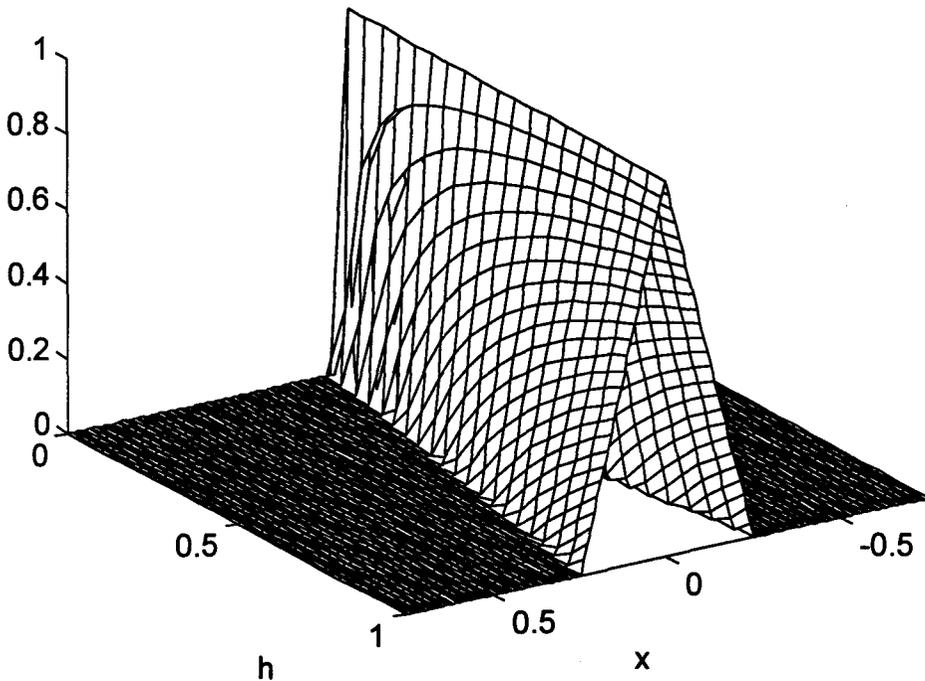


Fig. 2.28. Fusion avec le nouvel opérateur lorsque les deux sources sont absolument fiables. (les t_i sont constants égaux à 1). L'intersection est retrouvée.

De même il est possible de construire un pseudo opérateur de priorité, en donnant à π_1 une fiabilité certaine, et à π_2 une fiabilité égale à h . La figure 2.29 montre le résultat.

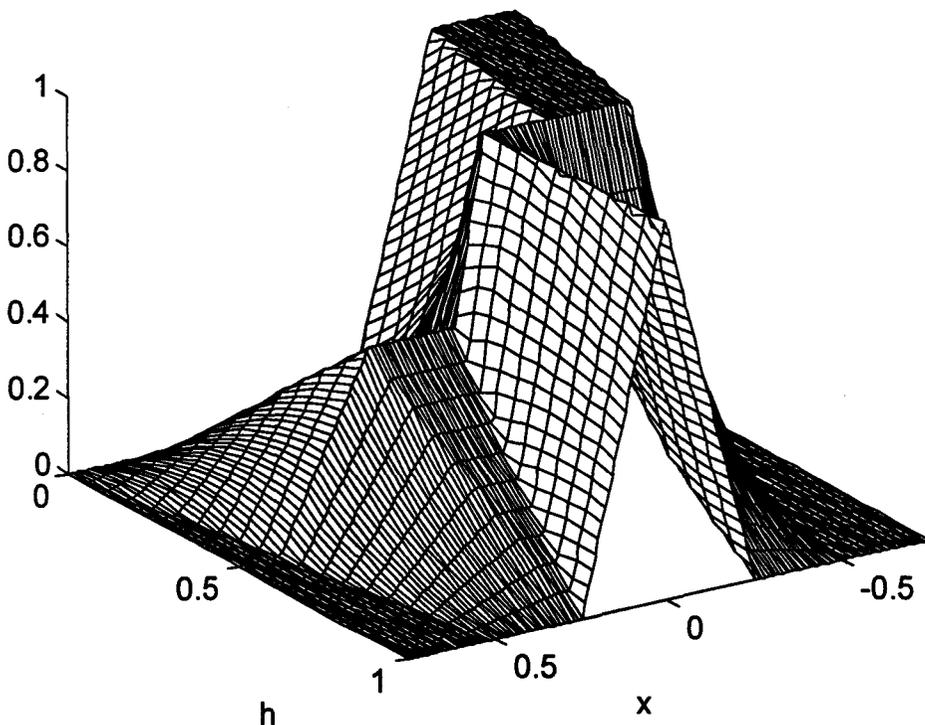


Fig. 2.29. Fusion avec pseudo priorité de π_1 sur π_2 .

Contrairement à l'opérateur de priorité, qui ne prend en compte les informations de π_2 que si elles sont en accord avec π_1 , dans ce résultat, l'ensemble des informations fournies par π_2 ont un degré de fiabilité non nul égal à h , ce qui explique que x proche de 0.5 a un degré de possibilité en général non nul (en fait non nul pour $h \neq 1$ et $h \neq 0$).

Il pourrait être envisagé de définir des coefficients de fiabilité flous permettant de donner un degré de fiabilité pour chaque information élémentaire $\pi(x)$, mais cette voie n'est qu'à peine explorée [Delmotte et al, 1996d].

Le nouvel opérateur, dans ses différentes versions, apparaît donc comme un opérateur englobant les principaux modes de fusion actuels. Cependant il nécessite la connaissance des degrés de fiabilité. Cette estimation, ou tout au moins une proposition d'estimation, est donné dans un prochain paragraphe, après avoir envisagé le degré de crédibilité qu'il est possible d'accorder au résultat.

2.4.3 Estimation de la crédibilité accordée au résultat.

Avec les opérateurs classiques d'agrégation, il n'y avait aucune estimation de la qualité du résultat. Par exemple avec l'intersection, le résultat était considéré bon, et il n'y avait que des méthodes subjectives et empiriques pour déterminer si le mode conjonctif était finalement pertinent ou non, et s'il fallait appliquer le mode disjonctif.

L'opérateur adaptatif de Dubois et Prade implante un mécanisme basé sur la quantité de conflit entre les sources pour passer du mode conjonctif au mode disjonctif, mais ensuite le résultat est encore une fois qualifié de correct et utilisé tel quel par la suite.

De même avec les méthodes de transformation des sources selon leurs degrés de confiance, les sources sont transformées puis jugées parfaitement fiables, et donc elles sont fusionnées par une méthode de type intersection. Le résultat est alors jugé pertinent et crédible.

Il est pourtant possible de construire deux séries de distributions différentes, avec deux séries de degrés de fiabilité différentes, telles que les deux fusions des deux séries de distributions donnent exactement la même information, avec le même degré de crédibilité, alors que les informations initiales ont des degrés de fiabilité différents.

En fait ceci s'explique parce que dans la théorie des possibilités, les données et leurs degrés de confiance sont confondus par l'intermédiaire d'une transformation qui traduit tout en termes de degrés de possibilités. Ce n'est pas parce qu'il est possible d'interpréter des distributions de possibilités très simples mathématiquement, comme des triangles avec une base plus ou moins haute par exemple, en termes de certitude de l'information et de précision des données qu'il faut identifier ces deux quantités. Bien au contraire il serait intéressant de séparer les deux informations, et d'associer à une distribution de possibilité donnée un certain degré de certitude ou de confiance. La nouvelle information serait alors le couple $(\pi(x),t)$ qui appartient à $X^{[0,1]} \times [0,1]$ ¹

Le nouvel opérateur de fusion présenté est un opérateur de $(X^{[0,1]} \times [0,1])^n \rightarrow X^{[0,1]}$ (n étant le nombre de sources d'informations). Il n'est donc pas un opérateur interne pour le nouvel espace $X^{[0,1]} \times [0,1]$.

¹ La notation $X^{[0,1]}$ désigne l'ensemble des distributions de possibilités pouvant être définies sur le référentiel X .

Il faudrait pouvoir lui ajouter un mécanisme estimant la fiabilité du résultat, de telle sorte que la nouvelle règle de fusion soit un opérateur interne. Ainsi dans des mécanismes de fusion itératifs, un résultat obtenu par une précédente fusion pourrait être agrégé immédiatement avec d'autres sources de données, pour obtenir une information globale de 2^{ème} ordre, et ainsi de suite.

La construction de ce mécanisme, tout au moins quelques orientations naturelles, est donnée dans les paragraphes qui suivent [Delmotte et al, 1996e].

2.4.3.1 Composantes de la fiabilité du résultat

Tout d'abord la crédibilité d'un résultat dépend des crédibilités accordées aux sources sur lesquelles il est basé. C'est le principe de la maxime 'pour les questions d'argent il vaut mieux aller voir son banquier qu'un dresseur de chien'. Lorsqu'un homme affronte un problème complexe, il cherche plutôt à interroger des experts, ou des membres de sa famille, deux types de sources auxquels il accorde une grande confiance.

Il est naturel qu'un résultat issu de la fusion de sources fiables soit plus fiable qu'un résultat basé sur des sources non fiables. Ce principe sera caractérisé par la confiance externe d'un résultat. Le mot externe est utilisé parce que cette composante ne fait pas référence au mode de fusion, et donc au résultat, mais est basée sur des informations qui ont été fournies, et qui sont donc externes à l'algorithme.

La deuxième composante dépend au contraire du résultat. Lorsque des informations contradictoires sont fusionnées, si le mode disjonctif est utilisé, le résultat est une union de différents modes, comme dans les exemples du paragraphe 2.4.2, où lorsque $h=0$, le résultat était l'union des deux trapèzes. Ce résultat était parfaitement ambigu, contrairement aux résultats donnés par un mode conjonctif. Si une décision devait être prise, quel événement $x \in [-0.75, 0.75]$ serait le plus important: un x proche de 0,5 ou proche de $x=-0.5$? Avec cette seule information il était impossible de trancher entre les deux cas.

La deuxième composante est donc basée sur le fait qu'un résultat est non ambigu, c'est à dire qu'il ne présente qu'un seul mode, ou encore que son cœur est un ensemble connexe, et non une union de différents pics de possibilités.

Cette caractéristique de non contradiction est à distinguer de la précision d'un résultat: le fait qu'un mode soit plus ou moins élargi ne change pas le fait que le cœur soit un fermé. Bien sûr le cœur contient en général plus d'un événement (comme les coeurs des trapèzes qui en contenaient une infinité), mais les événements d'un cœur connexe sont plus proches logiquement entre eux que des événements qui appartiennent à deux modes différents, même s'ils sont très proches. Cette deuxième composante est appelée contradiction interne.

Il faut maintenant fusionner ces deux quantités. Le choix retenu accorde plus d'importance à la fiabilité externe. Si des sources très peu fiables sont d'accord entre elles, le résultat sera non contradictoire, mais pourtant la crédibilité de ce résultat ne sera pas très élevée. Même si les sources sont d'accord elles peuvent se tromper. Pourtant le fait qu'elles soient d'accord entre elles, ceci étant mesuré par la deuxième composante, est bénéfique.

Aussi la deuxième composante sera utilisé comme une récompense, et donc sera utilisé comme un facteur correctif de la première.

Les paragraphes suivants exposent une méthode calculant la fiabilité externe, puis une méthode estimant la contradiction interne et finalement proposent une méthode de fusion.

2.4.3.2 Fiabilité externe

Cette composante est exclusivement basée sur les fiabilités des sources.

Plus les sources fusionnées sont fiables, plus le résultat est fiable, cette composante est donc une fonction croissante de ses arguments. Le nouvel opérateur de fusion est conçu de telle manière que les sources non fiables sont rejetées, au pire elles sont éliminées du résultat si leur fiabilité est nulle.

Dans ce sens les sources les plus fiables jouent un rôle prépondérant par rapport aux sources non fiables, et ceci doit se refléter dans la composante externe de fiabilité.

Aussi parmi la série des opérateurs possibles dans la théorie des possibilités, les t-normes sont à rejeter, car la fiabilité du résultat tendrait vers la plus basse des fiabilités.

Au contraire les t-conormes apparaissent beaucoup plus pertinentes. Les opérateurs de compromis, tels les moyennes, peuvent être envisagés, mais ne correspondent pas vraiment à la logique du nouvel opérateur.

Des exemples possibles sont:

$$t_m = \max(t_i), t_x = 1 - \prod(1 - t_i) \quad (2.54)$$

t_m utilise l'union de Zadeh et t_x l'union probabiliste.

Pour choisir entre ces deux mesures, quelques réflexions qui mettent en évidence le problème de renforcement des t-conormes sont proposées.

- Cas de la fusion de n sources également fiables, de fiabilité égale à $1/n$. Si le nombre de sources devient très grand, et même infini, quelle est la fiabilité du résultat?

Avec f_m le résultat aura une fiabilité nulle, avec f_x la fiabilité limite sera égale à $1 - 1/e$ (puisque $f_x = 1 - (1 - 1/n)^n$), qui sera une quantité non nulle, et même supérieure à 0.5.

- Cas de la fusion de n sources dont la fiabilité est égale à ϵ , avec $\epsilon \ll 1$.

Avec f_m la fiabilité du résultat est indépendante du nombre de sources, et est égale à ϵ .

Avec f_x la fiabilité limite lorsque n tend vers l'infini est égale à 1 ($f_x = 1 - (1 - \epsilon)^n$)

Dans les deux cas, bien que l'opérateur soit une t-conorme, les fiabilités finales apparaissent opposées. En fait avec l'opérateur de Zadeh, il n'y a pas de renforcement: le fait de fusionner plusieurs sources d'informations n'augmente pas la crédibilité finale par rapport à la crédibilité d'une source prise isolément.

Pour la suite l'union probabiliste a donc été retenue.

2.4.3.3 Contradiction interne

Cette mesure est basée sur le nombre de modes que présente un résultat. Plus un résultat présente de modes, plus la contradiction interne augmente, a fortiori s'ils sont éloignés les uns des autres

Etant fondée sur les différents modes d'un résultat, la forme de celui ci sera donc étudiée. Dans la suite, les différents modes d'une distribution de possibilité sont supposés connus. Par exemple pour un espace X unidimensionnel, une méthode de gradient en balayant tout l'espace peut être utilisée. En outre ils sont supposés classés et numérotés, l'indice p_i désignant le $i^{\text{ème}}$ mode, et les calculs nécessitent d'introduire la notion de distance entre ces modes.

Pour un problème défini sur un espace X multidimensionnel, il faudra d'abord classer les différents modes. Pour cela on construit un chemin minimal entre les centres des coeurs de chaque mode, qui ne passe qu'une fois par chaque mode, et qui minimise le chemin parcouru total (au sens de l'abscisse curviligne du chemin dans l'espace X). C'est le problème du voyageur de commerce qui se révèle très complexe lorsque le nombre de modes augmente. Il est supposé que ce problème est soluble.

Une distribution définie sur un espace X multidimensionnel est donc ramenée à un problème sur un espace X unidimensionnel. En choisissant une origine au chemin, il est possible de numéroté les modes, la distance entre eux étant simplement basée sur l'abscisse curviligne.

La figure 2.30 montre un exemple de distribution à 3 modes p_1, p_2 et p_3 .

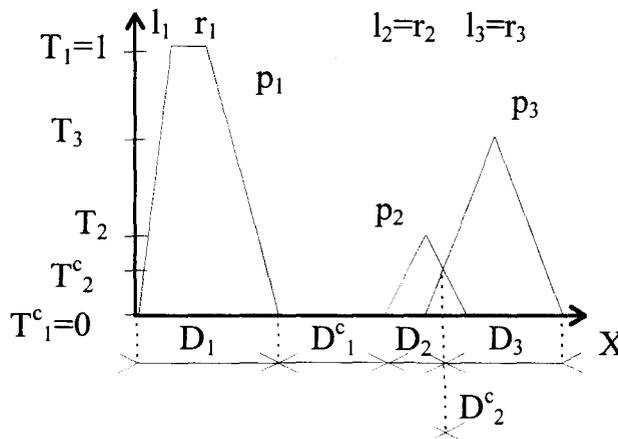


Fig. 2.30. Variables utilisées pour le calcul de la contradiction interne.

La distribution possède m modes p_i .

Le domaine D_i associé à un mode p_i est défini comme un ouvert: c'est l'ensemble des x tels qu'une méthode de gradient partant des ces points converge vers le maximum du mode.

Le maximum d'un mode, ou sa hauteur, noté T_i est défini par $T_i = \max_{x \in D_i} \pi(x)$ et le cœur du mode est noté $X_i = [l_i, r_i]$, et est calculé comme l'argument de T_i .

Le domaine D_i^c entre 2 modes successifs est défini comme suit:

$$D_i^c = [r_i, l_{i+1}] \cap (D_i \cup D_{i+1})^c$$

et la non contradiction entre deux modes est définie par

$$T_i^c = \max_{x \in D_i^c} \pi(x)$$

Ayant défini les variables utilisées pour calculer l'indice de contradiction (en fait ce sera un indice de non contradiction), il faut maintenant passer à sa construction.

De quelle quantité peut dépendre la quantité de conflit entre 2 modes successifs?

D'abord intervient la distance entre ces 2 modes: plus ils sont éloignés, plus la contradiction augmente.

Ensuite intervient la différence de hauteur entre ces 2 modes. Si 2 modes sont d'égale hauteur, le conflit est plus important que si un mode atteint une hauteur de 1, et que l'autre atteint une hauteur 10 fois plus petite.

Enfin le dernier facteur est l'éventuelle hauteur de l'intersection entre les 2 modes. S'il y a un point commun, le conflit est moindre que s'il n'y avait aucune intersection.

L'indice de contradiction entre 2 modes est donc le suivant:

$$\text{contr}_i = (r_i - l_{i+1}) \left(1 - |T_i - T_{i+1}|\right) \left(1 - \frac{T_i^c}{\max(T_i, T_{i+1})}\right)$$

Mais il est utilisé sous la forme d'un indice de non contradiction entre 2 modes successifs:

$$q_i = 1 - (r_i - l_{i+1}) \left(1 - |T_i - T_{i+1}|\right) \left(1 - \frac{T_i^c}{\max(T_i, T_{i+1})}\right) \quad (2.55)$$

Cet indice ne peut être calculé que si 2 modes ou plus existent.

Par définition l'indice de qualité d'une distribution unimodale est pris égal à 1.

Cette définition est compatible avec la définition 2.51.

En effet si deux modes sont examinés, et que l'un des deux disparaît, soit $(r_i - l_{i+1})$ tend vers 0 si les 2 modes se rapprochent jusqu'à se confondre, soit $(1 - |T_i - T_{i+1}|)$ tend vers 0 si un des modes s'effondre (note: tous ces calculs sont faits sur une distribution normalisée, donc au moins un mode atteint 1), soit encore $\left(1 - \frac{T_i^c}{\max(T_i, T_{i+1})}\right)$ atteint 0 si l'intersection augmente.

Il faut encore pondérer la quantité de non contradiction entre 2 modes. En effet plusieurs configurations de conflits peuvent donner le même indice de non contradiction, par exemple soit des modes dont la hauteur atteint 1, soit des modes dont la hauteur est proche de 0, or ces derniers sont moins gênants.

Ceci est donc réalisé par le coefficient w_i :

$$w_i = T_i + T_{i+1} \quad (2.56)$$

L'indice final de qualité du résultat est donc

$$q = \frac{1}{\sum_{i=1}^{m-1} w_i} \sum_{i=1}^{m-1} w_i q_i \quad (2.57)$$

Plus cet indice est proche de 1, plus la contradiction interne $1-q$ est faible.

2.4.3.4. Fusion de la fiabilité externe et de la contradiction interne

L'approche retenue est celle qui consiste à utiliser l'indice de qualité comme un facteur multiplicatif de la fiabilité externe. Les deux indices ne sont en effet pas de la même nature et il n'est pas possible de les fusionner directement, avec par exemple un opérateur de la théorie des possibilités.

Si la qualité du résultat est importante, le facteur multiplicatif sera supérieur à 1 (ce sera un bonus), et si la qualité du résultat est faible, le facteur multiplicatif sera inférieur à 1 (ce sera un malus). Le seuil de qualité pour la récompense est mis à $q=0.5$.

La relation générale est donc du type

$$t' = t(1 + g(t, q)) \quad (2.58)$$

avec les contraintes suivantes:

$$g : [0,1]^2 \rightarrow [0,1]$$

$$g(t, 0.5) = 0$$

$$g(t, q^+) \geq 0$$

$$g(t, q^-) \leq 0$$

$$g(t, q_1) < g(t, q_2) \text{ if } q_1 < q_2$$

pour $q^+ \in [0.5, 1]$, une grande qualité, et $q^- \in [0, 0.5]$, une mauvaise qualité.

Un grand nombre de fonctions est possible. Pour simplifier, la fonction g a été prise linéaire par morceau pour la variable q . g est en fait la réunion de deux segments de droites, un pour q^+ et un pour q^- . Afin d'assurer la continuité, les deux segments de droite se rejoignent en $q=0.5$. Les pentes des deux segments sont obtenues dans le cas limite où q^+ est maximal et vaut 1 et q^- minimal et vaut 0. Dans ces deux cas, le bonus et le malus ont été pris égaux à 50% de la fiabilité externe.

Afin d'obtenir une crédibilité finale directement normalisée, un paramètre supplémentaire vient saturer les deux droites, car un bonus de 50% appliqué à une fiabilité externe de 1 aurait donné sans cela une fiabilité finale de 1,5.

Les formules finales sont donc:

$$\begin{aligned} g(t, q^+) &= (2q^+ - 1) \min(0.5, 1 - t) \\ g(t, q^-) &= (2q^- - 1) \min(0.5, t) \end{aligned} \quad (2.59)$$

La figure 2.31 donne deux exemples de modification de la fiabilité externe en fonction de la non contradiction interne du résultat.

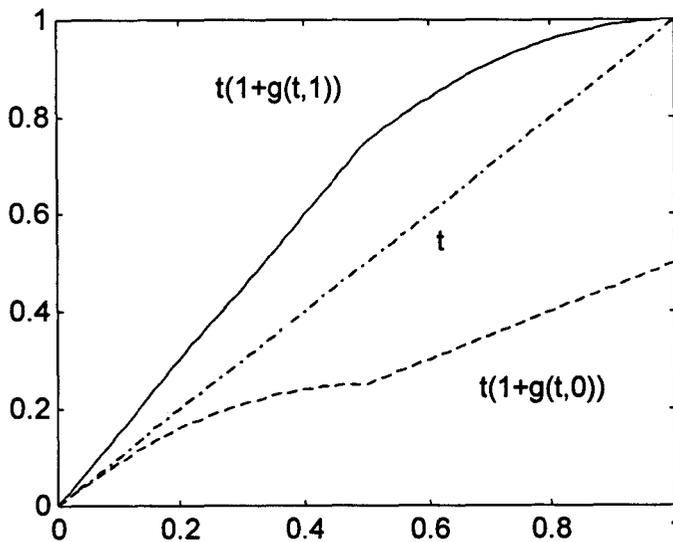


Fig. 2.31. Fiabilité finale en fonction de la fiabilité externe t et de deux indices de qualité.

Un résultat dont la fiabilité externe serait égale à 1 (les sources sont des experts parfaits) mais dont l'indice de qualité est nul (les sources sont en contradiction absolue) verrait sa fiabilité passer à 0.5.

2.4.4 Estimation des crédibilités

Dans ce paragraphe est donnée une méthode permettant d'estimer les coefficients de crédibilité de sources. En effet le nouvel opérateur nécessite la connaissance de ces coefficients pour être appliqué et il ne faut pas rejeter le problème du choix d'un opérateur sur le problème de l'estimation des coefficients de fiabilité. Résoudre des problèmes traitant de l'incertain en exigeant encore plus d'information serait alors assez paradoxal...

Avant de présenter l'approche développée, qui, en quelques mots, est basée sur le calcul d'indices de conflit entre les sources, est présenté l'approche habituellement utilisée dans un cadre probabiliste.

2.4.4.1 Calcul de la fiabilité dans un cadre probabiliste

L'approche présentée ici utilise toute une série de tests et donc un véritable apprentissage hors ligne des sources d'informations.

Dans le cadre probabiliste, la qualité d'un expert est le produit de 2 quantités: son informativité et sa calibration. L'informativité I d'une source mesure la précision des informations fournies. La calibration C mesure quant à elle l'exactitude des données apportées par une source d'information.

Pour calculer l'informativité I de la source s_i (représentée par une distribution de probabilité) est introduit d'abord la fonction mesurant la quantité d'information apportée par r alors que p est connue:

$$I(r, p) = \sum_{i=1}^n r(x_i) \ln(r(x_i) / p(x_i)) \quad (2.60)$$

dans un espace discret, et

$$I(r, p) = \int_{x_a}^{x_b} r(x) \ln(r(x) / p(x)) dx \quad (2.61)$$

dans un espace continu.

L'informativité d'un expert se calcule alors par rapport à la distribution de probabilité uniforme, parce que l'information a priori est l'ignorance totale, justement représentée par la distribution uniforme.

Par exemple dans le cas continu, l'informativité de la source s_i est:

$$I(s_i) = \ln(x_b - x_a) + \int_{x_a}^{x_b} s_i(x) \ln(s_i(x)) dx \quad (2.62)$$

Pour calculer la calibration d'un expert, la procédure est la suivante:

La source s_i est interrogée à propos d'un problème, ou mieux, à propos de m problèmes légèrement différents, et doit fournir un certain nombre de distributions de probabilités. Après avoir fait des moyennes, ces m distributions de probabilité sont comparées aux m véritables distributions de probabilités, obtenues par des séries de mesure. Pour cela est calculée la quantité d'information relative entre les 2 distributions de probabilités $s_{i,m}$ et p_m pour chaque problème posé.

Pour obtenir l'indice de calibration C de s_i , une fonction statistique est utilisée:

$$C(s_i) = 1 - \chi^2(2mwI(s_{i,m}, p_m)) \quad (2.63)$$

avec w un paramètre tel que $w \in [0.1, 1]$, et χ^2 la fonction utilisée en statistique.

L'indice final de qualité d'une source s_i est alors $Q(s_i) = I(s_i)C(s_i)$, qui signifie qu'un expert très précis mais qui se trompe toujours aura une qualité nulle, et inversement une source qui est très proche de la vérité parce que trop peu précise aura aussi une qualité très faible.

Cette approche est donc rigoureuse, mais nécessite beaucoup de données concernant les sources d'informations, ce qui n'est pas toujours le cas. Une version de cette méthode a été proposée pour les possibilités [Sandri et al, 1995].

2.4.4.2 Objectif de l'algorithme d'estimation des crédibilités

L'objectif de l'algorithme présenté dans ce paragraphe est de faire une estimation des degrés de confiance dans les sources sans devoir passer par un quelconque apprentissage. L'algorithme devra pouvoir être appliqué sans autres données que les informations à fusionner.

Quelles sont les informations disponibles? L'exemple qui suit propose quelques orientations.

Soit un capteur mesurant une variable. A ce capteur est associé une série de conditions d'utilisation, par exemple des conditions de température, d'humidité, et si ces conditions ne sont pas respectées, le capteur pourra d'abord fournir des informations légèrement faussées, puis complètement fausses, et à mesure que les conditions d'utilisation différeront des conditions conseillées, la crédibilité qui sera accordée aux informations fournies par le capteur diminuera. De même, aux capteurs sont associés des temps moyens de fonctionnement avant

panne. Si un capteur a un temps d'utilisation prévu de 20000 heures, et qu'il est utilisé depuis 30000 heures, la confiance qui lui est accordée diminuera.

Par ailleurs il est possible que ce capteur ne soit pas le seul à mesurer une variable donnée. Ainsi il pourrait y avoir une redondance dans les informations utilisées pour estimer la validité des informations fournies. Par exemple en surveillance les informations fournies par un capteur sont toujours vérifiées par d'autres capteurs, afin de pallier une de leurs défaillances possibles. Si trois capteurs sont en jeu, et que brusquement un des trois capteurs se met à donner des mesures erratiques ou en contradiction avec les deux autres qui continuent à être en accord, un problème sera sûrement apparu.

Il apparaît donc que les informations disponibles peuvent être classées en deux groupes.

La première concerne les informations extérieures aux données fournies par les sources. Par exemple la comparaison des conditions d'utilisation conseillées et réelles pour un capteur entre dans ce groupe. Il est supposé que les informations de cette classe sont directement accessibles et modélisées par des degrés de confiance, appelés confiances externes car dépendants de données autres que les sources d'informations.

L'algorithme doit pouvoir intégrer de telles données, mais elles ne sont pas obligatoires. Dans la suite, de telles informations sont supposées données sous la forme du coefficient $t_{ext} \in [0,1]$.

Le deuxième groupe d'informations relatives à l'estimation des confiances correspond aux comparaisons des sources entre elles. Cette classe forme le cœur de l'algorithme et est détaillée dans le prochain paragraphe. Un degré de crédibilité noté t_{int} du même type que la crédibilité externe sera alors calculée, basé cette fois sur les sources. Il faudra finalement fusionner les deux indices.

2.4.4.3 Estimation de la fiabilité interne

Cette partie est basée sur la comparaison des sources entre elles pour déterminer les éventuels désaccords. La méthode sera celle suivie en analyse de données avec l'analyse hiérarchique. Cette branche de l'analyse de données classe des populations d'éléments, appelés individus, grâce à leur différences ou leur ressemblances.

Ces classements utilisent des indices de ressemblance ou de similarité entre les individus, qui sont rappelés ci dessous.

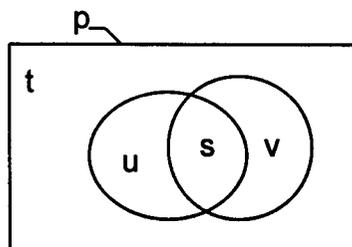


Fig. 2.32. Cas de figure pour les différences entre individus.

Chaque individu correspond à p variables binaires différentes, et ils sont comparés 2 par 2. Lorsqu'ils sont comparés, l'ensemble s des variables pour lesquels les deux individus ont une valeur égale à 1 est défini, de même l'ensemble t des variables pour lesquels les valeurs sont nulles. Les ensembles u et v caractérisent les variables telles que leur valeur soit égale à 1 pour l'individu i et 0 pour l'individu j, et réciproquement pour l'ensemble v.

Une fonction r_{ij} comparant les individus i et j est un indice de ressemblance si cette fonction est croissante pour le paramètre s et décroissante pour les paramètres u et v (symétriquement, et donc puisque $p=s+t+u+v$, r_{ij} doit être croissante pour le paramètre t) [Lerman, 1970].

Des indices classiques sont:

$$\text{-Russel et Rao} \in [0,1] \quad r_{ij} = s/p \quad (2.64)$$

$$\text{-Kendall} \in [0,1] \quad r_{ij} = 1 - (u + v) / p \quad (2.65)$$

$$\text{-Jackard} \in [0,1] \quad r_{ij} = s / (s + u + v) \quad (2.66)$$

$$\text{-Pearson} \in [-1,1] \quad r_{ij} = (st - uv) / \sqrt{(s+u)(s+t)(t+u)(t+v)} \quad (2.67)$$

Parmi les indices de ressemblance sont définis les indices de similarité, qui ont la particularité d'être des fonctions constantes lorsque $u=v=\emptyset$. Ils permettent d'avoir des échelles communes de comparaison. Le premier indice n'est pas un indice de similarité.

Tous ces indices ont des comportements différents qu'il serait trop long d'exposer ici. Par exemple l'indice de Pearson provient d'un indice de corrélation et celui de Russel et Rao est le plus simple. Une remarque importante est qu'ils ne sont pas en général équivalents, à moins d'introduire la contrainte que les partitions des p valeurs de chaque individu en valuation 1 et en valuation 0 doivent avoir exactement le même cardinal. Par exemple si $p=3$, et qu'un individu soit $\{1,0,0\}$, tous les autres individus devront contenir exactement un 1 et deux 0. Dans le cas contraire, ils donnent naissance à des ordres différents.

Ces indices peuvent être étendus au cas de variables floues dans le cas continu (le référentiel X a les bornes a et b). Les trois premiers indices deviennent ainsi:

$$r_{ij} = \frac{\int_a^b \cap(\pi_i, \pi_j) dx}{b - a} \quad (2.68)$$

$$r_{ij} = \frac{\int_a^b \cap(\pi_i, \pi_j) dx + \int_a^b (1 - \cup(\pi_i, \pi_j)) dx}{b - a} \quad (2.69)$$

$$r_{ij} = \frac{\int_a^b \cap(\pi_i, \pi_j) dx}{\int_a^b \cup(\pi_i, \pi_j) dx} \quad (2.70)$$

Si ces indices étaient utilisés pour comparer les sources, les temps de calculs seraient très long. Pour en augmenter la rapidité, un indice très frustre est utilisé à leur place:

$$r_{ij} = \max_x \cap(\pi_i, \pi_j) \quad (2.71)$$

Cet indice est appelé indice de compatibilité entre π_i et π_j . C'est en fait la classique hauteur h de l'intersection entre les deux distributions.

Cet indice permet donc la construction d'une matrice symétrique de dimension $n \times n$, reprenant toutes les valeurs des indices de compatibilité entre les sources. En analyse de données cette

matrice est l'objet de nombreuses études, par exemple pour tenter de réduire le nombre de sources s'il apparaît que certaines sont dépendantes d'autres.

Cette matrice est utilisée pour calculer les indices de fiabilité interne $t_{i,int}$:

$$t_{i,int} = \frac{\sum_{j \neq i} r_{ij}}{n-1} \quad (2.72)$$

La relation 2.72 est une simple somme normalisée. D'autres opérateurs peuvent être éventuellement utilisés, mais dans ce cas il faut utiliser des opérateurs de type compromis. En effet les opérateurs logiques ne conviennent pas: l'intersection prendra le plus petit indice de conflit, c'est à dire que dès qu'une source sera en contradiction avec n-1 autres, toutes auront de très faibles fiabilités. De même l'union ne convient pas, pour des raisons inverses: il suffit qu'une source soit en accord avec une autre pour avoir un très haut indice de fiabilité.

La relation 2.72 implante l'opérateur de compromis le plus simple qui soit.

Pour qu'une source ait un coefficient de fiabilité interne nul, il faut qu'elle soit en contradiction absolue avec toutes les autres sources, et inversement pour qu'une source ait une fiabilité interne égale à 1, il faut qu'elle soit en accord avec toutes les autres.

2.4.4.4 Fusion des fiabilités externes et internes.

L'indice 2.71 estimant la compatibilité entre deux sources est très simple, mais il ne prend pas en compte toute l'information disponible: ce n'est pas un opérateur parfait, et il peut induire en erreur l'utilisateur. Pour résumer il est possible de dire que la fiabilité des informations obtenues par cet opérateur n'est pas absolue.

De même il se peut que les informations externes estimant la fiabilité externe soient fausses ou trompeuses, mais qu'il ne soit pas possible d'estimer la fiabilité des informations externes fournies.

Les opérateurs de type logique n'apparaissent donc pas adéquats pour fusionner les deux indices de fiabilité, comme le souligne l'exemple suivant:

Un nouveau capteur ($t_{ext}=1$) est en contradiction avec une série d'autres, plus âgés ($t_{int}=\epsilon$). Apparemment le nouveau capteur est défectueux. Mais peut être sont-ce les anciens qui sont tous défectueux. Suivant l'hypothèse retenue, l'opérateur logique devrait être soit une intersection, soit une union. Si l'introduction d'une hypothèse gêne, il peut être prudent d'opter pour un opérateur de compromis.

La relation 2.73 propose un opérateur de compromis le plus simple, en donnant autant d'importance aux deux indices, externe et interne.

$$t = 0.5(t_{int} + t_{ext}) \quad (2.73)$$

L'algorithme estimant les degrés de fiabilité des sources est donc complètement défini par cette dernière formule. Il peut intégrer des informations externes aux sources, mais peut aussi donner une évaluation des crédibilités sans autre information. Il a été utilisé dans les simulations pour tester les différents modes de fusion, lorsque le mode de fusion avait besoin

de coefficients de fiabilité. L'intersection entre les deux trapèzes avait en effet une hauteur h (qui est utilisée dans la relation 2.71), comme il n'y avait que deux modes, la relation 2.72 n'était pas appliquée et 2.71 donnait directement la fiabilité des deux trapèzes.

Le paragraphe qui suit donne un exemple d'opération qui peut être effectué sur les degrés de crédibilité, indépendamment des données, puisque maintenant les deux variables sont séparées.

2.4.5 Optimisation des degrés de crédibilité

La méthode d'estimation des degrés de crédibilité des sources utilise des indices de compatibilité ou de ressemblance. Aucun de ceux exposés n'est meilleur qu'un autre, en tout cas pas d'une manière absolue, et chacun d'eux a ses propres caractéristiques.

Il n'est pas possible de définir un indice universel (déjà faudrait il caractériser le sens d'universalité) qui puisse englober tous les autres indices et qui ait toutes les propriétés souhaitables. Depuis longtemps en analyse de données ce sujet est un thème de recherche, et il n'a pas encore abouti. En outre l'indice de conflit utilisé réellement (par la relation 2.71) est très simple, et ne prend pas en compte toute l'information disponible.

La conclusion de ces remarques est que l'algorithme d'estimation des crédibilités n'est pas parfait et qu'il peut éventuellement faire des erreurs sur les estimations des degrés de confiance.

Or ces degrés de fiabilité donnés aux sources d'informations sont d'une importance capitale pour le nouvel opérateur de fusion. Ainsi les figures 2.28 et 2.29 diffèrent-elles totalement de la 2.27 car les coefficients de fiabilité sont différents, bien que les sources soient elles-mêmes identiques.

Pour des systèmes d'aide à la décision il existe des décisions à prendre qui se révèlent plus importantes que d'autres. En outre il existe de plus en plus des systèmes d'aide à la décision basés sur la théorie des possibilités.

Dans de tels cas il pourrait être intéressant de développer un mécanisme, éventuellement coûteux en temps de calcul, mais qui assurerait d'obtenir des degrés de confiance ou de fiabilités "très proches de la réalité", afin de pouvoir effectuer des fusions de données les plus fiables et les plus précises qui soient. C'est l'objectif du paragraphe suivant, qui propose un mécanisme d'optimisation des coefficients de crédibilité, afin de pallier aux insuffisances des estimateurs basés sur des indices de conflits [Delmotte et al, 1997a].

2.4.5.1 Construction du problème d'optimisation

Lorsque le mot optimisation est prononcé, il faut toujours préciser pour quel critère il l'est. En effet une solution optimale fait toujours référence à la fonction de coût ou au critère qu'il a fallu, selon les circonstances, maximiser ou minimiser.

Quel pourrait être le critère d'optimisation pour les fiabilités?

En fait avec la version $\pi_{x_{or}}$ avant normalisation, si les sources ne sont pas fiables, ou si elles sont en contradiction, le résultat tend vers une distribution de possibilité constante égale à 0.

Et inversement si les sources sont fiables et en accord, le résultat est presque normalisé et sa hauteur est très proche de 1.

Aussi un critère possible d'optimisation pourrait être la hauteur du résultat avant normalisation.

Quels sont les paramètres à optimiser? Il s'agit des coefficients de fiabilité. Pourtant le problème d'optimisation des fiabilités ne doit pas chercher à calculer précisément chacun des degrés de crédibilité t_i indépendamment de toute donnée précédente. Non seulement cela demanderait un temps de calcul énorme, mais en plus cela serait absurde: il existerait alors en effet un grand nombre d'optimum globaux, tous en contradiction (par exemple $t_1=1$, et tous les autres nuls, ou encore $t_2=1$ et tous les autres nuls...A chaque fois le résultat est déjà normalisé, pourtant dans le premier cas seule la première source est utilisée, dans le second cas seule la deuxième l'est).

Bien au contraire il faut prendre en compte les informations précédentes (c'est à dire les coefficients de fiabilité estimés), qui ne sont peut être pas parfaits mais relativement proches de la réalité. L'algorithme sera simplement utilisé pour corriger les valeurs estimées afin de maximiser la hauteur du résultat avant normalisation.

Si les valeurs estimées n'étaient pas prise en compte, jamais l'algorithme d'optimisation avec son seul critère ne pourrait converger vers les véritables valeurs.

Il reste avant de définir le problème d'optimisation à préciser comment seront liés les paramètres d'optimisation et les paramètres de crédibilité estimée.

En fait les coefficients de fiabilité estimée seront légèrement corrigés par des paramètres additifs. Afin de simplifier le problème, comme avec une hypothèse d'équiprobabilité en probabilité, une hypothèse 'd'equicorrection' est formulée: en fait il y a un seul paramètre θ à optimiser, le problème étant finalement de maximiser la fonction:

$$f(\theta) = \max_x \left(\text{Xor}(t_i - \theta) \max[(t_i - \theta)\pi_i(x)] + \text{And}(t_i - \theta) \min[(t_i - \theta)\pi_i(x)] \right) \quad (2.74)$$

avec $\theta \in [-1, 1]$

Ainsi les coefficients de crédibilité sont légèrement ajustés, d'une manière identique pour tous, de telle sorte que le résultat final soit le plus haut possible avant normalisation, ce qui indique le résultat le plus fiable possible.

Dans le cas de la figure 2.28, où les fiabilités étaient constantes égales à 1, le résultat avant normalisation était très proche de zéro lorsque les deux trapèzes étaient en contradiction (h proche de 0). Les fiabilités fournies étaient alors complètement fausses.

L'algorithme d'optimisation détectera immédiatement le problème et ajustera les fiabilités de manière à les diminuer (θ positif) de telle sorte que le résultat soit un peu plus proche d'un mode disjonctif.

La relation 2.74 peut être simplifiée, tout au moins l'espace de recherche peut être réduit, avec une hypothèse supplémentaire. Il est en effet possible de supposer que l'information pertinente fournie par l'estimateur de crédibilité soit seulement l'ordre relatif des fiabilités des sources, c'est à dire que l'estimateur réalise un classement légèrement quantifié.

Il est alors possible de supposer qu'au moins une des sources est absolument fiable.

Une transformation simple est alors appliquée aux degrés de confiance estimés:

$$t_i' = t_i + \beta \text{ avec } \beta = 1 - \max(t_i)$$

Ainsi seuls les écarts de crédibilités sont conservés. Par cette transformation, des problèmes au départ différents peuvent donner des vecteurs de crédibilité transformés identiques.

Comme les coefficients de crédibilité sont par définition des nombres $t_i \in [0,1]$, il est alors possible de réduire l'espace de recherche:

$$f(\theta) = \max_x \left(\text{Xor}(t_i' - \theta) \max[(t_i' - \theta)\pi_i(x)] + \text{And}(t_i' - \theta) \min[(t_i' - \theta)\pi_i(x)] \right) \quad (2.75)$$

avec $\theta \in [0,1]$.

Une fiabilité inférieure à 0 sera prise dans l'algorithme de recherche égale à 0.

Le critère est identique pour π_{Nand} , mais pas pour π_{Nor} . En effet avec cette version, il existe au moins un optimum global correspondant à $\theta=1$, mais dans ce cas le résultat est l'ignorance complète. Il faut donc pour π_{Nor} minimiser aussi la variation de crédibilité. Par exemple

$$f(\theta) = \max_x \left(\text{Xor}(t_i' - \theta) \max[(t_i' - \theta)\pi_i(x)] + \text{And}(t_i' - \theta) \min[(t_i' - \theta)\pi_i(x)] + \text{Nor}(t_i) \right) - k\theta \quad (2.76)$$

avec k une pondération.

Dans tous les cas le problème est de trouver

$$\theta_{\text{opt}} = \text{Arg}(\max_{\theta} f(\theta)) \quad (2.77)$$

2.4.5.2 Résolution du problème

Bien qu'il n'y ait qu'un paramètre de recherche, le problème est complexe et la fonction $f(\theta)$ a en général des maximums locaux. L'exemple suivant (construit exprès) montre ce problème.

Quatre sources de données π_1, π_2, π_3 et π_4 sont fusionnées. Ce sont de simples trapèzes comme à la figure 2.8.

En fait π_1 a un degré de fiabilité estimé maximal, égal à 1, mais se retrouve en contradiction avec les trois autres, qui sont d'accord entre elles, mais ont des fiabilités estimées moindres.

Les données sont les suivantes:

π_1 a une fiabilité de 1, π_2 a une fiabilité de 0.73, avec une intersection de 0.2 avec π_1 . π_3 a une fiabilité de 0.7 et une intersection de 0.25 avec π_1 . π_4 est égale à π_3 .

La figure 2.33 montre les évolutions de $f(\theta)$ obtenue par 2.75 avec les quatres sources.

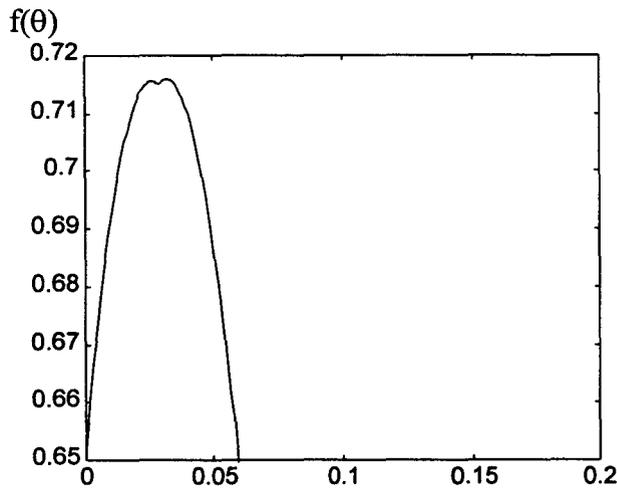


Fig. 2.33. Maximum de la fusion avant normalisation lorsque les fiabilités sont adaptées.

Le premier maximum atteint en partant de $\theta = 0$ n'est pas le maximum global, aussi des méthodes de descente ne conviennent pas en général.

Cet exemple est intéressant parce qu'il montre que les méthodes de fusion basées sur des coefficients de fiabilité sont des fonctions sensibles de ces derniers. En effet le premier maximum atteint correspond à des x appartenant au cœur de π_1 , tandis que le second maximum est atteint pour des x appartenants au cœur de π_2 . Le maximum de possibilité saute brusquement entre les deux coeurs, et un algorithme d'aide à la décision aurait très bien pu proposer les deux choix possibles sans pouvoir trancher.

La conclusion de ce paragraphe est que le problème d'optimisation est complexe, et que la fonction à optimiser a en général plusieurs maximums locaux.

Il faut donc se tourner vers des algorithmes d'optimisation multi-modes, qui sont bien plus complexes que des algorithmes de type gradient. Pour la simulation du paragraphe suivant, les algorithmes génétiques furent utilisés. Des notions concernant ces derniers sont donnés en annexe.

2.4.5.3 Exemple d'optimisation

Le problème est la fusion des deux classiques trapèzes. Le paramètre h varie et règle le niveau de conflit entre les sources. En outre un paramètre supplémentaire est utilisé: la confiance en la deuxième source. En effet pour cette simulation, la confiance en la première source est prise constante et égale à 1, et la confiance en la deuxième source est prise comme un paramètre supplémentaire qui varie aussi entre 0 et 1.

Les paramètres de l'algorithme génétique sont les suivants:

θ est codé sur 7 bits, ce qui correspond à un précision de 1%, la population est de 20 individus sur 10 générations. La probabilité de cross over est égale à 0.7 et le taux de mutation est pris égal à 0.01. En outre un élitisme de 1 individu est appliqué à chaque génération, ainsi qu'un rééchelonnement linéaire des valeurs du critère lors de l'étape de sélection.

La figure 2.34 donne la valeur optimale du paramètre θ de correction de fiabilité.

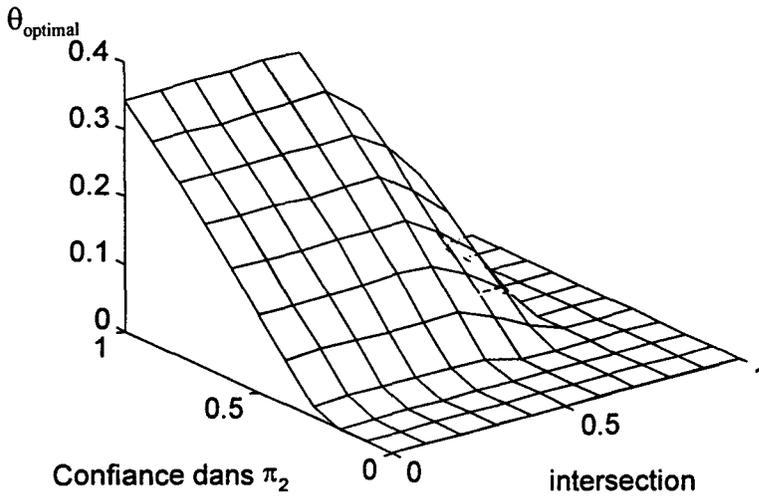


Fig. 2.34. Décroissance optimale de fiabilité.

Les résultats de la fusion après optimisation sont des courbes en 4D ($X, \pi(x)$, le paramètre h de l'intersection et la confiance dans la seconde source), qui ne peuvent pas être très facilement imprimés. Des coupes seront données à la place.

La figure 2.34 montre quel paramètre est prépondérant lors d'un conflit. Lorsque deux distributions sont fusionnées, la première servant de référence, est-ce la crédibilité dans la seconde source qui est le paramètre important, ou le niveau réel de conflit entre les deux sources?

La figure 2.34 permet de répondre à cette question puisqu'elle montre les valeurs optimales de décroissance de crédibilité lorsqu'un conflit est détecté. Plus cette décroissance est importante, plus le conflit est sévère.

Or l'analyse de la figure 2.34 montre que c'est le paramètre de la confiance qui est prépondérant. En effet dès que la confiance dans la seconde source décroît, le conflit après l'application de la fusion diminue, comme le montre la décroissance du paramètre θ . Par contre, le conflit ne diminue pour le paramètre de l'intersection h qu'à partir du moment où l'intersection dépasse 0.5 (mais alors il est vrai que la décroissance du paramètre θ est très rapide).

Dans les cas où il n'y a pas de conflit, comme par exemple lorsque la seconde source a une fiabilité nulle, le résultat est déjà optimal au sens du critère défini, et il n'y a pas d'optimisation possible ($\theta = 0$).

Les figures suivantes donnent les résultats avant optimisation et après (note: l'axe indicé par h est équivalent à l'intersection).

Les figures 2.35 et 2.36 donnent les résultats avec une confiance dans la source 2 constante et égale à 1.

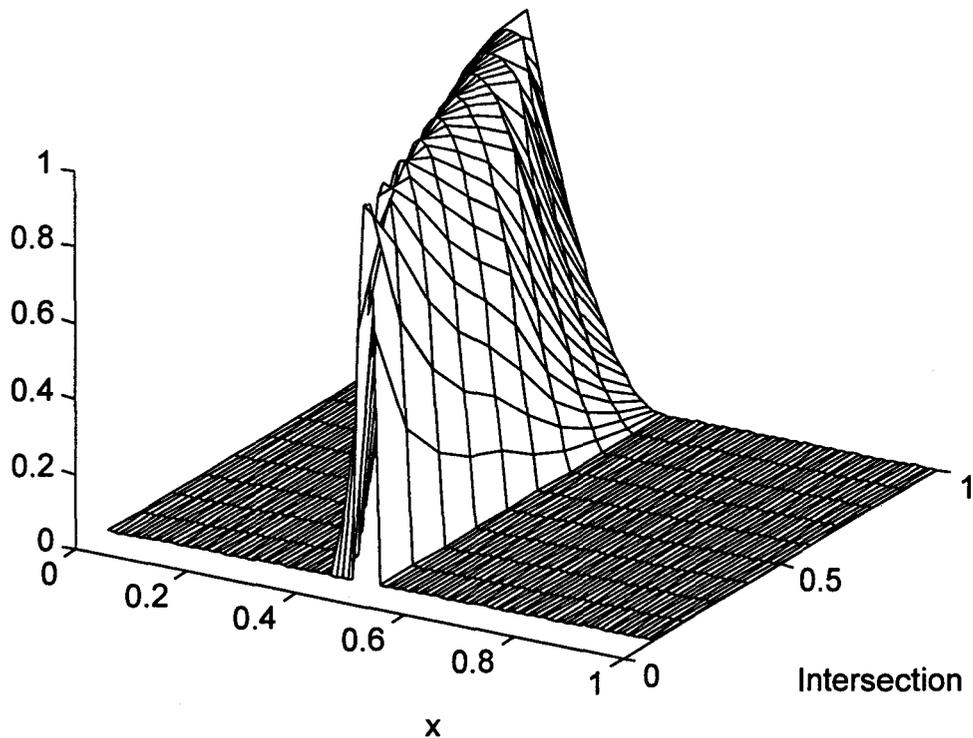


Fig. 2.35. Résultat de la fusion avant optimisation ($t_2 = 1$).

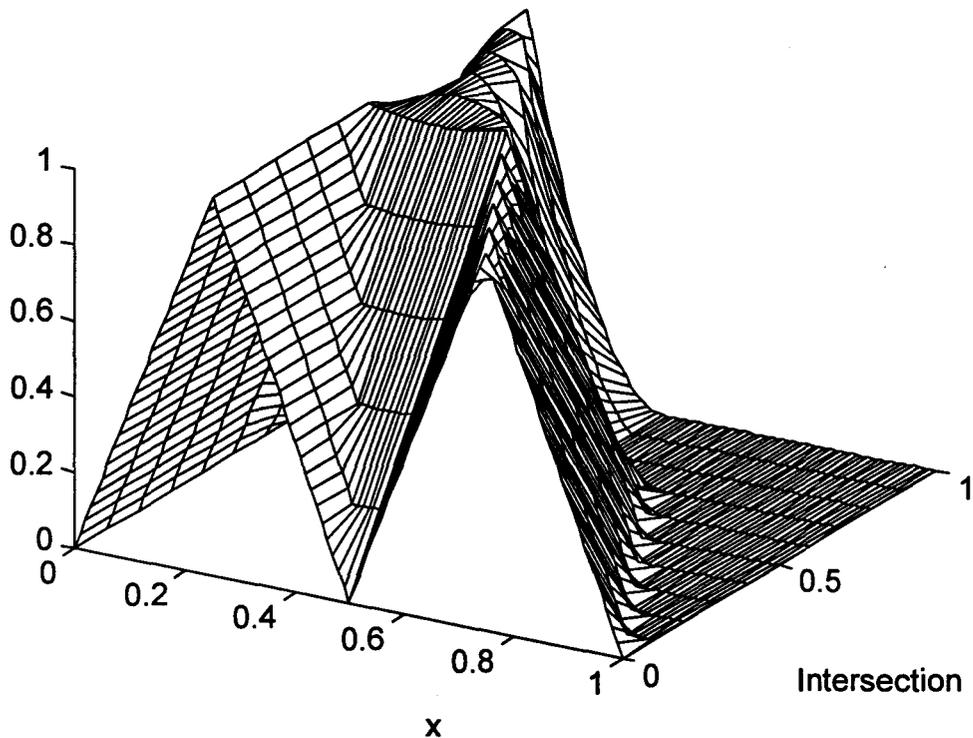


Fig. 2.36. Résultat de la fusion après optimisation ($t_2 = 1$).

Les informations utilisées pour obtenir 2.35 sont contradictoires pour $h = 0$. En effet les deux sources sont supposées être absolument fiables, or leur intersection est vide ($h = 0$). Avec une intersection classique (le nouvel opérateur est en effet une intersection dans ce cas), le résultat n'est pas défini. A mesure que les deux sources se rapprochent, l'intersection augmente, et le conflit diminue. Pour $h = 1$, l'intersection est parfaitement retrouvée.

L'optimisation va détecter immédiatement le conflit pour h proche de 0. Les confiances sont adaptées, en fait diminuées, et le résultat est en fait l'union. Cette union tend alors vers l'intersection lorsque le conflit diminue.

Les figures 2.37 et 2.38 donnent les résultats pour une coupe avec une intersection constante égale à 0.5.

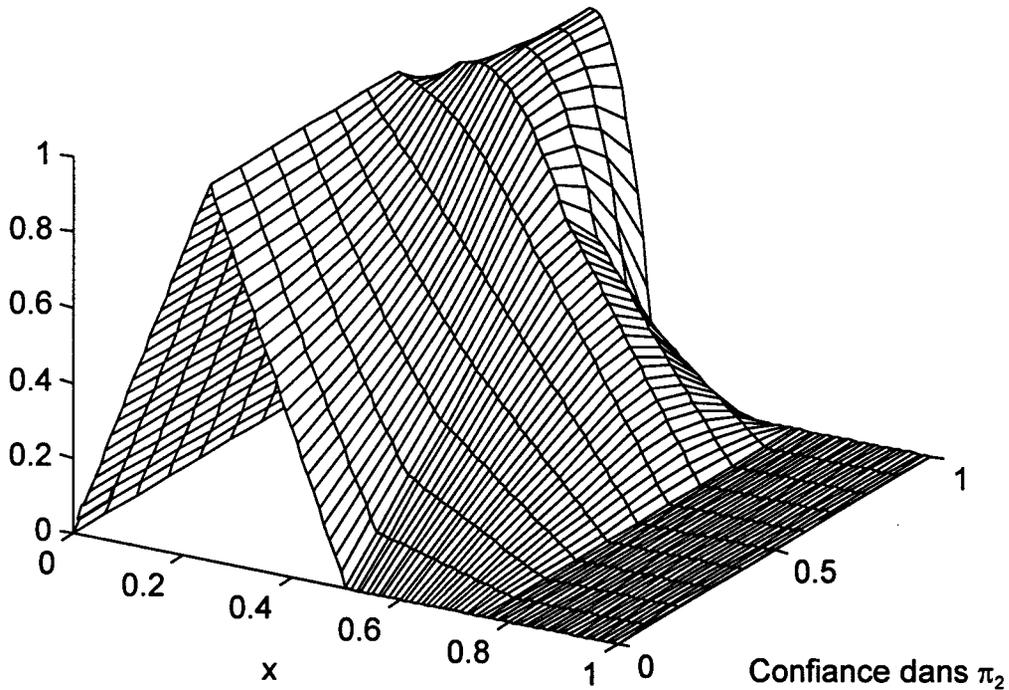


Fig. 2.37. Résultat de la fusion avant optimisation ($h = 0.5$).

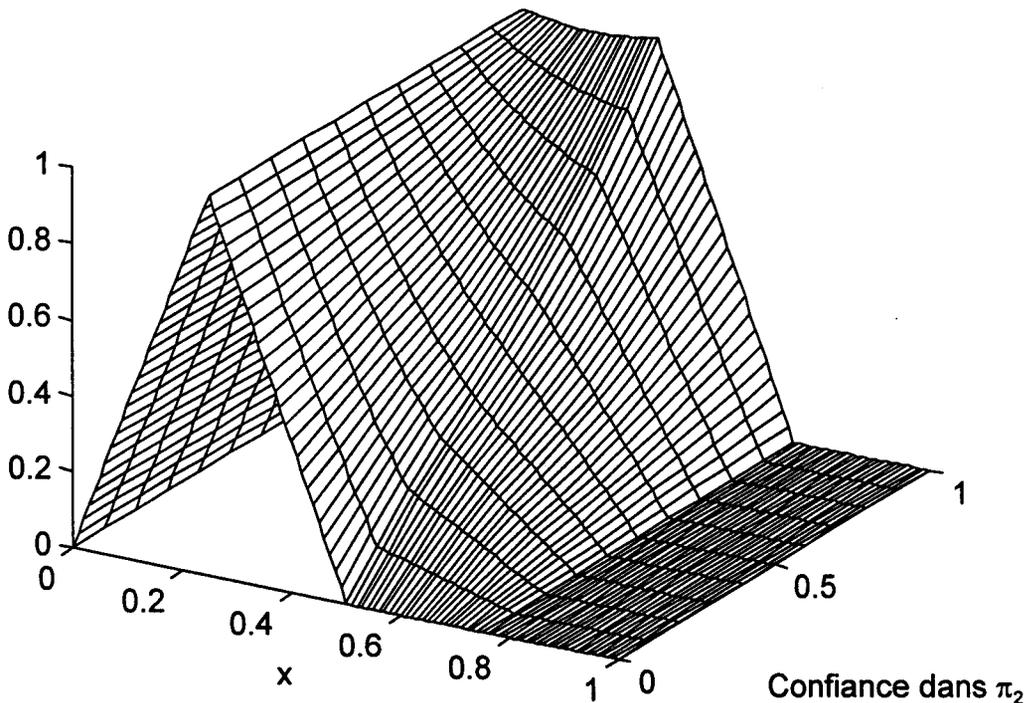


Fig. 2.38. Résultat de la fusion après optimisation ($h = 0.5$).

Avec cet exemple le conflit n'est pas total, puisque l'intersection atteint 0.5. La figure 2.37 montre une préférence pour une fusion conjonctive, alors qu'avec l'optimisation le résultat est un peu plus prudent, et que la fusion de type disjonctive est renforcée.

2.4.6 Exemple final: le paradigme du policier contre les voleurs

Cet exemple a pour but de montrer comment s'agencent tous les outils définis avec le nouvel opérateur, et ce type de problème, par les variables qu'il met en jeu, illustre bien le problème de la fusion de données. Enfin il ne laisse personne indifférent.

Un crime a été commis par une personne X. Un policier a été témoin du crime et affirme que la personne B est le coupable. Cependant la personne B a un alibi, puisque n-1 de ses amis affirment que B était bien avec eux lorsque le crime fut commis. Comme le titre de l'exemple l'indique, ces amis sont des personnes de fiabilité douteuse, par rapport au policier assermenté.

Que doit penser le juge de ces informations, en dehors de tout indice matériel? Obligatoirement un des deux camps se trompe ou ment, mais lequel choisir?

Les outils développés avec le nouvel opérateur aideront le juge à prendre sa décision.

D'abord l'univers de discours X doit être défini. Cet ensemble X est très simple, puisque il est binaire: $X = \{\text{oui}, \text{non}\}$, oui étant compris comme 'la personne B est bien coupable'.

Il y a n+1 sources en jeu: le policier, la personne B et ses n-1 amis, qui forment en fait 2 types de sources distinctes:

Il y d'une part la source π_p , modélisée par $\pi_p(\{\text{oui}\})=1$ et $\pi_p(\{\text{non}\})=0$;
et d'autre part les n sources π_{r_i} qui affirment que B est innocent, modélisées par $\pi_{r_i}(\{\text{oui}\})=0$ et $\pi_{r_i}(\{\text{non}\})=1$.

Ces deux types de sources sont en conflit absolu.

Il faut maintenant envisager les fiabilités de ces sources.

Le policier est assermenté, donc sa fiabilité externe est prise égale à 1: $t_p^{\text{ext}} = 1$

Les amis de la personne B, et la personne B, sont des personnes bien connues de la police, ce sont toutes des personnes associées au grand banditisme.

Aussi leur fiabilité externe est-elle très faible: $t_r^{\text{ext}} = \varepsilon$ avec $\varepsilon \ll 1$.

Il faut maintenant estimer leur fiabilité interne.

Celle du policier est nulle, car il est en contradiction complète avec toutes les autres sources.

En effet son indice de ressemblance $r(p,r_i) = 0 \forall i \in \{1,n\}$ et donc $t_p^{\text{int}} = 0/n = 0 \forall i \in \{1,n\}$

Par contre les n voleurs sont tous d'accord entre eux, aussi $r(r_i,r_j) = 1 \forall (i,j) \in \{1,n\}^2$ et donc $t_{r_i}^{\text{int}} = (n-1)/n \forall i \in \{1,n\}$

En fusionnant les deux fiabilités externes et internes, les fiabilités finales sont égales à:

$$t_p = 0.5$$

$$t_p = 0.5$$

$t_{r_i} = 0.5(1+c) \quad \forall i \in \{1, n\}$ avec $c = \varepsilon - 1/n$, qui peut être positif ou négatif suivant les valeurs de ε et de n .

La structure générale du résultat obtenu avec la version π_{xor} est

$$\pi = f \max(t_p \pi_p, t_{r_i} \pi_{r_i})$$

En effet le terme conjonctif est nul, puisque les $n+1$ sources sont dans leur ensemble en contradiction totale.

f est un facteur de normalisation qui tient compte du coefficient Xor, mais il ne change pas le résultat.

A cause de la structure des $n+1$ sources de distributions, les deux degrés de possibilités sont:

$$\pi(\text{oui}) / f = t_p \quad \text{et} \quad \pi(\text{non}) / f = t_{r_i}$$

Aussi le degré maximal de possibilité avant normalisation sera soit atteint pour oui soit atteint pour non, selon que $t_{r_i} = 0.5(1+c)$ est inférieur ou supérieur à $t_p = 0.5$

La conclusion est donc que le juge croira le policier si $c < 0$, c'est à dire si le nombre de voleurs est inférieur à $1/\varepsilon$.

Il reste maintenant à envisager la confiance que le juge accordera à son analyse.

D'abord la fiabilité externe est donnée par une union probabiliste:

$$f_x = 1 - 0.5[1 - 0.5(1+c)]^n = 1 - 0.5^{n+1}(1-c)^n$$

C'est une fonction croissante qui tend vers 1 lorsque le nombre de voleurs tend vers l'infini (ceci est du au phénomène de renforcement).

Maintenant il faut évaluer l'indice de qualité ou de non contradiction.

La distribution de possibilité ne possède que 2 modes correspondant à $x=\text{oui}$ et $x=\text{non}$.

La distance entre ces 2 modes est comprise comme étant la plus grande possible puisque ces 2 propositions représentent des sens absolument opposés.

De même l'intersection entre ces 2 modes est nulle.

En fait le seul facteur changeant dans la contradiction est la hauteur relative des 2 modes, et finalement d'indice de qualité vaut:

$$q = 1 - \left(1 - |t_{pn} - t_{r,n}|\right)$$

avec

$$t_{pn} = \begin{cases} 1 & \text{si } c < 0 \\ 1 / (1+c) & \text{sinon} \end{cases}$$

et

$$t_{r,n} = \begin{cases} 1+c & \text{si } c < 0 \\ 1 & \text{sinon} \end{cases}$$

En simplifiant les expressions, l'index de qualité vaut finalement:

$$q = \begin{cases} -c & \text{si } c < 0 \\ c / (1 + c) & \text{si } c \geq 0 \end{cases}$$

Cette expression dépend à la fois du nombre de voleurs et de la fiabilité externe ϵ accordée aux voleurs. Il est minimal lorsque $c=0$, c'est à dire lorsque le résultat présente 2 modes strictement égaux.

Enfin, grâce à cet index de qualité, la crédibilité finale peut être calculée:

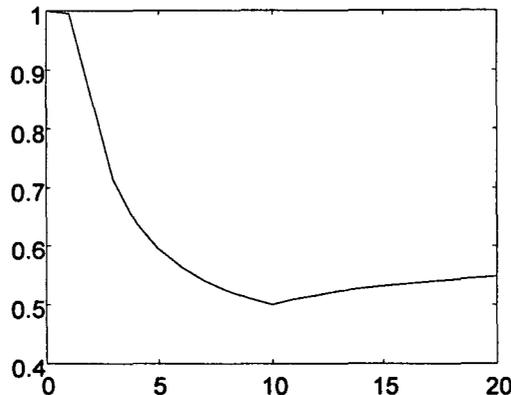


Fig. 2.39. Crédibilité finale du juge dans son analyse, en fonction du nombre de voleurs. La crédibilité externe ϵ des voleurs est égale à 0.1.

Lorsque le nombre de voleurs est nul (la personne B ne proteste pas), le juge croit le policier et sa crédibilité dans son analyse est maximale. Puis à mesure que le nombre d'amis de B augmente, le juge croit encore le policier, mais il commence à douter, puisque son sentiment de confiance dans le résultat chute.

Puis il arrive que le juge soit dans l'incertitude la plus complète, lorsque le nombre de voleurs égale $1/\epsilon$ (dans l'exemple 10).

Après le juge change d'opinion et croit en l'innocence de B, mais sa crédibilité dans le résultat est faible et n'augmente que faiblement, puisque le policier, lui, est toujours présent pour affirmer le contraire.

La crédibilité du juge tend vers $0.5 + \epsilon / (1 + \epsilon)$, c'est à dire que pour $\epsilon \ll 1$, elle sera toujours proche de 0.5.

Cet exemple, parce que touchant à la justice, a provoqué de vives controverses. Une attitude commune était de dire qu'au bénéfice du doute, dès que B avait un ami témoignant pour lui, il fallait le relâcher, parce que le policier pouvait se tromper. A fortiori si B avait 15 personnes pour témoigner en sa faveur.

Mais dans ce cas il n'y avait pas de réponse de la part des contradicteurs au cas, plus complexe, où il y avait 2 policiers qui accusaient B et ses amis, et aux cas encore plus complexes où x policiers affrontaient y voleurs.

Cet exemple illustre une proposition qui tend à résoudre les conflits en prenant en compte et en les modélisant les degrés de crédibilité accordés aux sources d'informations.

L'estimateur fourni par les relations 2.71 à 2.73 apparaît comme étant un estimateur additif. Littéralement le policier 'pèse' autant que $1/\varepsilon$ voleurs. Il serait possible de définir d'autres estimateurs non additifs, mais en l'occurrence c'est le seul qui fut développé.

2.5 Conclusion: vers une théorie des possibilités étendue?

Ce chapitre a exposé les principes de la théorie des possibilités pour représenter et traiter les données incertaines.

Les informations peuvent être imprécises, incomplètes ou non fiables, et il apparaît que les mesures de possibilités sont plus dédiées à la gestion des données imprécises.

Pour réduire cette imprécision, des opérateurs de fusion ont été développés afin de pouvoir agréger des sources de données différentes, en vue de corroborer ou d'infirmer les données actuelles.

Certains des opérateurs développés permettent de prendre en compte un indice de fiabilité des sources, mais dans les seules approches n'impliquant pas de contrainte sur ces indices, les indices de fiabilité sont intégrés au moyen de transformations directement dans les données elles-mêmes.

Ceci est du au fait que dans la théorie des possibilités des formes très simples de distributions sont interprétées en terme de précision et de certitude. Pourtant pour des distributions un peu plus complexes, comme celle de la figure 2.40, cette dichotomie n'est plus possible, et il ne reste plus que des informations de même nature, en l'occurrence des degrés de possibilités.

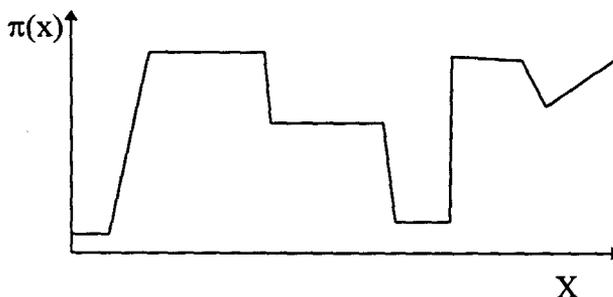


Fig. 2.40. Exemple de distribution non interprétable en terme de précision/certitude.

Pourtant il semble que les informations sous forme de degrés de possibilité et les informations de crédibilité qui sont accordées à ces degrés de possibilité soient réellement différentes.

Le nouvel opérateur est une proposition qui permet de garder distinctes les deux notions. A chaque distribution est associée un degré de confiance, et lors du processus de fusion si ces derniers sont utilisés pour choisir tel ou tel mode de fusion et pour pondérer l'influence de telle ou telle source d'information, ils ne sont pourtant pas confondus.

C'est pourquoi un opérateur de fusion des degrés de crédibilité, qui reste bien sûr lié au processus de fusion, a été développé. Ainsi à l'issue du processus de fusion une nouvelle information est générée, ainsi que son degré de fiabilité.

Cette nouvelle méthode de fusion pourrait servir de point de départ pour étendre la théorie des possibilités, où aux distributions de possibilités seraient associées des degrés de crédibilité. Les mesures de possibilités et de nécessités ne seraient pas modifiées, seulement leur seraient ajoutés des degrés de crédibilités. Il faudrait étudier les mesures de possibilités conditionnelles et la révision des possibilités, et peut être quelques autres mesures pour arriver à cet objectif de séparation définitive entre les données et leur crédibilités.

Ceci pourrait aussi être envisagé dans la théorie des fonctions de croyance où ce problème apparaît aussi, bien que sous des formes plus complexes.

Pour la suite le nouvel opérateur de fusion sera utilisé comme un des outils de l'analyse multi-modèle. En effet le multi-modèle est constitué de briques élémentaires qui sont ensuite fusionnées pour générer une information globale, ou une commande globale si le multi-modèle est utilisé en vue de la synthèse d'un contrôleur.

