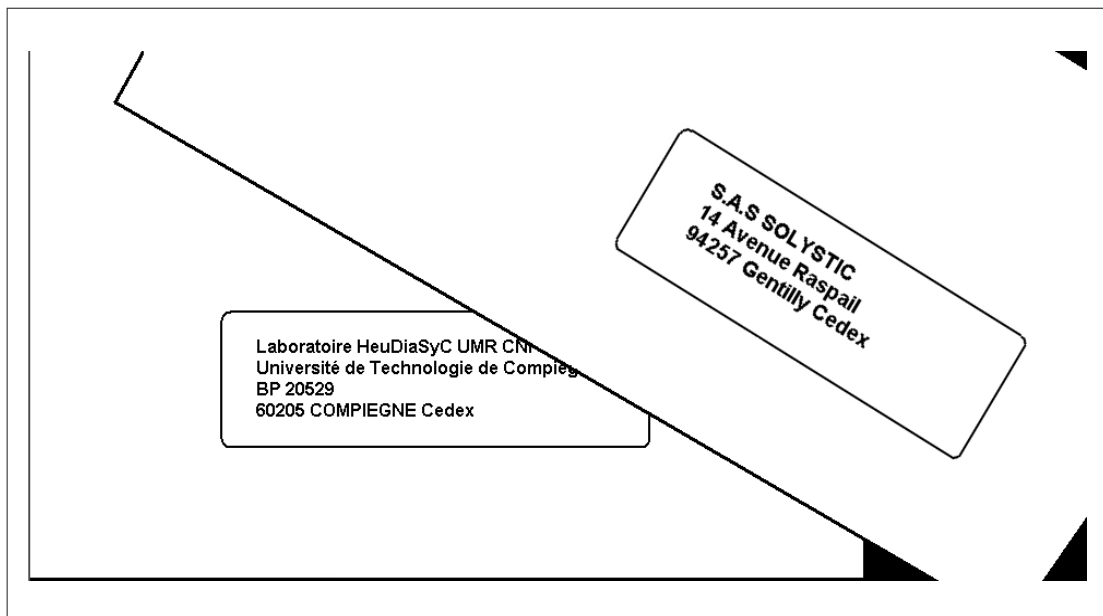


par David Mercier

Fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance

Thèse présentée pour l'obtention du grade de Docteur de l'UTC.



Soutenue le : 6 décembre 2006
Spécialité : Technologie de l'information et des systèmes

Fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance

Thèse soutenue le 6 décembre 2006 devant le jury composé de :

(Président)	B. Dubuisson	Professeur à l'Université de Technologie de Compiègne.
(Rapporteur)	I. Bloch	Professeur à l'École Nationale Supérieure des Télécommunications, Paris.
(Rapporteur)	M. Rombaut	Professeur à l'Université Joseph Fourier, Grenoble.
(Examineur)	B. Benyoub	Chef du département OVS, Solystic, Gentilly.
(Examineur)	G. Cron	Ingénieur de recherche, Solystic, Gentilly.
(Directeur de thèse)	T. Denœux	Professeur à l'Université de Technologie de Compiègne.
(Directeur de thèse)	M-H. Masson	Maître de conférences (HDR) à l'Université de Picardie Jules Verne, Amiens.

« Ignorance is precious, for once lost, it can never be regained. »- Dubois, Prade et Smets [39].

« Le subjectif est justement ce réel qu'on cherche à modéliser de façon rigoureuse. »- D. Dubois, extrait d'une conférence autour de l'aide à la décision et de l'intelligence artificielle, congrès AFCET, juin 1993.

« La probabilità non esiste di per sé. »- B. de Finetti [17]

Remerciements

J'aimerais témoigner ici de mes profonds remerciements à tous ceux qui de différentes façons m'ont aidé pour mener à bien les travaux présentés dans ce mémoire.

À mes directeurs de thèse, Thierry Denceux et Marie-Hélène Masson, pour avoir accepté de m'encadrer sur ce travail, pour leur suivi, leur écoute, leurs conseils, leur disponibilité. Ils ont su me laisser une certaine liberté et me reprendre toujours avec justesse et précision quand il le fallait. Je désire leur exprimer ici ma profonde gratitude.

À mon encadrant à Solystic, Geneviève Cron, ingénieur de recherche, pour sa formation et son aide à l'utilisation des outils Solystic, pour sa collaboration lors de la réalisation et l'implémentation de ce travail. Celui-ci a grandement bénéficié de son aide notamment en ce qui concerne les méthodes d'affectation de fonctions de croyance à partir d'une matrice de confusion. Enfin, par sa bonne humeur et son dynamisme, travailler avec elle fut toujours un plaisir.

Aux membres du jury :

Madame Isabelle Bloch et Mademoiselle Michèle Rombaut, pour avoir accepté de rapporter sur ce travail, pour les corrections apportées à ce mémoire, ainsi que pour toutes leurs remarques et suggestions.

Monsieur Belkacem Benyoub, chef du département OVS, pour avoir accepté ma candidature sur ce travail, pour toute son aide dans les affaires administratives, pour ces conseils et pour ces discussions enrichissantes sur le métier du postal.

Monsieur Bernard Dubuisson, pour m'avoir fait l'honneur de présider ce jury.

Un grand merci aussi :

À Stéphane Poirier, ingénieur de recherche à Solystic, responsable du pôle de compétence manuscrit, pour les nombreuses discussions que nous avons eues sur la reconnaissance du courrier postale ainsi que sur la combinaison.

À toute l'équipe OVS de Solystic, pour la bonne humeur générale, l'ambiance et la solidarité entre les personnes. La liste des noms est bien longue, mais ceux-ci se reconnaîtront : ils sont tous, au jour de la rédaction de cette page, au troisième étage et un peu au deuxième étage du bâtiment Solystic à Gentilly. Un merci tout spécial à Monsieur Christian Gigot, ingénieur de recherche à Solystic, pour ses célèbres organisations de pots et déjeuners sur l'herbe, mais aussi pour son aide précieuse dans le traitement du courrier postal. Un grand merci à tous les membres des équipes de natation et de course à pied de Solystic, et un grand bravo à Messieurs Hervé Perrin et Eric Buissonnet pour l'organisation. J'avoue ici être très souvent aller le mercredi à Gentilly, rien que pour faire coïncider mon travail à Solystic, et le rendez-vous piscine de la Butte aux Cailles.

À Benjamin Quost, collègue doctorant, compatriote de bureau depuis trois ans, pour toutes ces discussions que nous avons eues, et notamment pour sa collaboration lors du travail sur l'affaiblissement contextuel. Je remercie aussi Stéphane Aubry, doctorant et nouveau collègue de bureau, non seulement pour son partage du savoir nippon mais aussi pour toute sa sympathie. Je remercie de même François Clautiaux et David Savourey, doctorants, pour leur accueil chaleureux et leur aide. Merci aussi à Astride Aregui, Amel Ben-Yaghlane et Frédéric Pichon, doctorants au laboratoire HeuDiasyC, à Emmanuel Ramasso, doctorant au Laboratoire des Images et des Signaux de Grenoble, et Véronique Cherfaoui, Maître de conférences à l'Université de Technologie de Compiègne pour toutes les discussions sur les fonctions de croyance et pour tous les transferts d'informations. J'adresse aussi mes remerciements à tous les membres du thème ASTRID, à tous les autres doctorants gi-motivés, ainsi qu'à tout le personnel du secrétariat du laboratoire Heudiasyc pour leur disponibilité et aide.

À mes directeurs de DEA, Leila Amgoud, chargé de recherche CNRS à l'Institut de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier (UPS), et Andréas Herzig, directeur de recherche CNRS à l'IRIT, UPS, pour leurs soutiens et encouragements à poursuivre mon chemin dans la recherche.

À Monsieur Alexander Zwonkin, Professeur à l'Université de Bordeaux 1, pour ses qualités de pédagogue, pour son souci du partage du savoir et de la connaissance, des relations humaines, et pour sa contribution à me guider vers la recherche.

Enfin, un grand merci à mes parents, à tous les membres de ma famille, notamment à mes cousins picards que j'ai retrouvés à mon arrivée à Compiègne, à Franck et Aurélie ainsi qu'à toute ma co-famille Nono, Gord, Raph, Nath, Julie, Émilie, Sophie.

Table des matières

Remerciements	vii
Table des matières	ix
Notations	xiii
Liste des tableaux	xv
Liste des figures	xvii
Introduction	1
1 Introduction au domaine postal	5
1.1 Concept d'adresse postale et définitions	5
1.1.1 Structure générale d'une adresse postale	5
1.1.2 Structure détaillée des adresses postales françaises	7
1.2 Lecteurs d'adresses postales (LAP)	8
1.2.1 Des décisions multi-niveaux	8
1.2.2 Types d'écriture d'un courrier	9
1.2.3 Modélisation générique d'un LAP	10
1.3 Problématique de la fusion des décisions postales	11
1.4 Conclusion	13
2 Modèle des croyances transférables	15
2.1 Introduction	15
2.2 Représentation des connaissances	16
2.2.1 Concepts de base	16
2.2.2 Définitions	17
2.2.3 Représentation matricielle	18
2.3 Affectation d'une fonction de croyance par matrice de confusion	20
2.3.1 Matrice de confusion	20
2.3.2 Affectation Bayésienne	21
2.3.3 Affectation de Xu et al. [120]	22
2.3.4 Une affectation basée sur le taux de fiabilité	23
2.4 Manipulation des connaissances	24
2.4.1 Révision versus mise à jour	24
2.4.2 Le principe de minimum d'information	24

2.4.3	L'idée de transfert des masses	25
2.4.4	Spécialisation et généralisation	25
2.4.5	Connaissances distinctes	27
2.4.6	Conditionnement et déconditionnement	27
2.4.7	Combinaisons conjonctive et disjonctive	29
2.4.8	Affaiblissement	31
2.4.9	Le degré de conflit $m(\emptyset)$	32
2.4.10	Grossissements et raffinements	33
2.5	Niveau pignistique	34
2.5.1	Transformation pignistique	34
2.5.2	Cadre de pari	34
2.5.3	Risques pignistique, inférieur et supérieur	35
2.6	Conclusion	36
3	Fusion de décisions postales dans le cadre du MCT	37
3.1	Choix du cadre de discernement	37
3.2	Modèles hiérarchiques d'affectation	42
3.2.1	Affectation hiérarchique	42
3.2.2	Affectation hiérarchique catégorisée	45
3.2.3	Affectation postale employée	49
3.2.4	Combinaison des fonctions de croyance	51
3.3	Prise de décision	52
3.4	Résultats expérimentaux	55
3.4.1	Choix des coûts	55
3.4.2	Schémas à base de votes	56
3.4.3	Phase de test	58
3.4.4	Remarques	58
3.5	Conclusion	63
4	Mécanismes de correction de fonction de croyance	65
4.1	Introduction	65
4.2	Extension vide et déconditionnement dans le cas d'un espace produit	66
4.2.1	Marginalisation et extension vide dans le cas d'un espace produit	67
4.2.2	Conditionnement et déconditionnement dans le cas d'un espace produit	67
4.3	Affaiblissement classique	68
4.4	Affaiblissement contextuel	71
4.4.1	Expression de la fiabilité	71
4.4.2	Expression de αm	72
4.4.3	Discussion	74
4.4.4	Cas particuliers et exemples	75
4.5	Affaiblissement contextuel basé sur un grossissement	78
4.5.1	Expression de la fiabilité	79
4.5.2	Expression de αm	79
4.6	Apprentissage automatique des coefficients d'affaiblissement	82
4.6.1	Apprentissage de la fiabilité d'un capteur seul	83

4.6.2	Apprentissage de la fiabilité d'un groupe de capteurs	85
4.7	Mécanismes généraux de correction	86
4.7.1	Renforcement de Dencœux et Smets [29]	86
4.7.2	Schéma étendu de Zhu et Basir [127]	87
4.7.3	Expression générale d'un mécanisme de correction	88
4.8	Conclusion	91
5	Améliorations du modèle	93
5.1	Utilisation de l'affaiblissement contextuel	93
5.2	Prise en compte de scores de confiance	96
5.3	Influence de la dépendance des LAP	108
5.4	Prise en compte de décisions filtrées	112
5.5	Conclusion	112
	Conclusion générale et perspectives	115
	Bibliographie	117
A	Preuves des énoncés du chapitre 4	129
A.1	Preuves des propositions 4.1 et 4.2	129
A.2	Modification des croyances sur la fiabilité	131
A.3	Preuves des propositions 4.5 et 4.6	132
A.4	Preuve de la proposition 4.10	134

Notations

Généralités

x	Variable d'intérêt.
$\Omega = \cup_{k=1}^K \{\omega_k\}$	Ensemble fini des valeurs pouvant être prise par x , appelé, cadre de discernement ou univers de discours.
2^Ω	Ensemble des parties de Ω .
A, B	Des sous-ensembles de Ω .
$ A $	Cardinal de A , c'est-à-dire nombre d'éléments de Ω présents dans A .
\bar{A}	Complémentaire de A dans Ω .
$\Theta = \cup_{\ell=1}^L \{\theta_\ell\}$	Un cadre de discernement.
Ag, S	Un agent rationnel, une source d'information.
$[0, 1]$	Ensemble des nombres réels compris entre 0 et 1.
$\llbracket 1, N \rrbracket$	Ensemble des entiers naturels compris entre 1 et N .
A_*, A^*	Réductions intérieure et extérieure de A , définies à partir d'un grossissement de Ω .

Fonctions de croyance

m	Une fonction de masse.
b	Une fonction d'implicabilité.
q	Une fonction de communalité.
pl	Une fonction de plausibilité.
bel	Une fonction de croyance.
$m_{Ag,t}^\Omega[EC]\{x\}$	Notation complète, due à Smets, d'une fonction de masse tenue par un agent Ag , relative à la valeur prise par x dans Ω , au temps t , étant donné un ensemble de connaissances EC .
\mathcal{M}^Ω	Ensemble des fonctions de masse définies sur 2^Ω .
$\mathbf{m}, \mathbf{b}, \mathbf{q}, \mathbf{pl}, \mathbf{bel}$	Vecteur de masse (resp. implicabilité, communalité, plausibilité, croyance) associé à m (resp. b, q, pl, bel).

m_B	Fonction de masse catégorique sur $B \subseteq \Omega$.
m_Ω	Fonction de masse vide.
\bar{m}	Négation de m .
${}^{tr}m$	Fonction de masse m non vide, totalement renforcée.
$m^{\Omega \uparrow \Omega \times \Theta}$	Extension vide de m^Ω sur l'espace produit $\Omega \times \Theta$.
$m^\Omega[D] \uparrow \Omega \times \Theta$	Déconditionnement de $m^\Omega[D]$ sur l'espace produit $\Omega \times \Theta$.
α	Taux d'affaiblissement, un scalaire entre 0 et 1.
β	Degré de croyance dans la fiabilité d'une source, égal à $1 - \alpha$.
$\boldsymbol{\alpha}$	Vecteur d'affaiblissement, dont les composantes sont des taux d'affaiblissement α_i .
${}^\alpha m$	Fonction de masse m affaiblie par un taux d'affaiblissement α .
${}^\alpha m$	Fonction de masse m contextuellement affaiblie par un vecteur d'affaiblissement $\boldsymbol{\alpha}$.
${}^\alpha \mathbf{G}$	Matrice de généralisation associée à l'opération d'affaiblissement par un taux d'affaiblissement α .
${}^\alpha \mathbf{G}$	Matrice de généralisation associée à l'opération d'affaiblissement contextuel de vecteur d'affaiblissement $\boldsymbol{\alpha}$.
γm	Fonction de masse m corrigée par des paramètres $(\gamma_i, m_i)_i$, définie au paragraphe 4.7 du chapitre 4.

Prise de décision

\mathcal{D}	Ensemble de décisions possibles.
$BetP$	Probabilité pignistique.
$c : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$	Une fonction de coût, $c(d, \omega)$ représente le coût de décider d sachant que la vérité est ω .
Γ	Cadre de Pari.
ρ_{bet}	Risque pignistique.

Postal

LAP	Acronyme de Lecteur d'Adresses Postales.
OCR	Acronyme d'« Optical Character Recognition », employé à la place de LAP dans le milieu professionnel, désigne aussi un composant d'un LAP en charge de la reconnaissance de caractères.
A_i	Un lieu d'acheminement.
B_j	Une distribution spéciale.
R_k	Une voie.

Liste des tableaux

1.1	Exemples d'adresses françaises de même acheminement, mais de distributions géographiques différentes.	5
1.2	Exemple d'adresses françaises comportant des distributions spéciales.	6
1.3	Exemples d'adresses traitées respectivement au Royaume-Uni, en Hollande et au Brésil.	6
1.4	Performances moyennes réalisées par les LAP Solystic en fonction du type de courrier et du niveau.	12
2.1	Ordre des éléments d'un vecteur de masses associé à une fonction de masse définie sur $\Omega = \{a, b, c\}$	19
2.2	Illustration d'une matrice de confusion.	21
2.3	Exemple d'une matrice de confusion d'un classifieur en charge de la reconnaissance des caractères numériques de 1 à 9. Il est possible d'observer 60 décisions correctes, 16 décisions en erreur, et 20 rejets.	21
3.1	Un exemple de matrice de confusion associée à un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 3.5.	44
3.2	Matrice de confusion d'un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 3.6.	47
3.3	Matrice de confusion associée à un LAP sur la base de données réduite de l'exemple 3.1. Notations : cat = catégorie, bp = boîtes postales, vnn = voies non numérotées, pp = pas de portes, vn = voies numérotées, ad = acheminement distribuable, and = acheminement non distribuable.	50
3.4	Coûts de décisions en fonction de la décision fournie et de la vérité.	54
4.1	Matrice de généralisation associée à l'affaiblissement contextuel dans le cas $K = 2$	75
4.2	Matrice de généralisation associée à l'affaiblissement contextuel dans le cas $K = 3$	75
4.3	Matrice de généralisation associée à l'affaiblissement Θ -contextuel dans le cas $K = 3$ et $\Theta = \{\{\omega_1\}, \{\omega_2, \omega_3\}\}$	82
4.4	Données de l'exemple 4.5 issues de [40].	85

Liste des figures

1	Schémas de combinaison de trois lecteurs d'adresses postales.	2
1.1	Niveaux de décisions autorisées pour chaque LAP avec des exemples associés.	9
1.2	Exemples de courriers dactylographiés et manuscrits.	10
1.3	Modélisation d'un lecteur d'adresses postales.	11
1.4	Représentation graphique des performances à un certain niveau de quatre LAP.	13
2.1	Représentation des éléments focaux de certaines classes de fonctions de croyance.	18
2.2	Illustration [27] d'un grossissement Θ de Ω , avec raffinement ρ de Θ . . .	34
3.1	Des exemples de lettres dont l'adresse est invalide.	38
3.2	Rappel de la hiérarchie de référence.	39
3.3	Éléments du cadre de discernement associés à l'exemple 3.1.	40
3.4	Hiérarchie associée à l'exemple 3.1 et lien avec la hiérarchie de référence.	41
3.5	Exemple d'une hiérarchie à trois niveaux.	43
3.6	Un exemple de hiérarchie à trois niveaux avec des catégories de décisions.	46
3.7	Éléments du cadre de pari Γ construit à partir de Ω , de la hiérarchie de référence (figure 3.2), et des masses associées à l'exemple 3.8.	53
3.8	Performances des LAP et de différents points de fonctionnement de la combinaison CMCT réalisées sur l' ensemble d'apprentissage	57
3.9	Performances des LAP, des schémas à base de votes majoritaires et des points de fonctionnement de la combinaison CMCT sélectionnés, réalisées sur l' ensemble de test	59
3.10	Nombres de bonnes reconnaissances et d'erreurs pour chaque LAP et chaque réglage de la combinaison sur le lot de courrier d'apprentissage .	61
3.11	Nombres de bonnes reconnaissances et d'erreurs pour chaque LAP et chaque réglage de la combinaison sur le lot de courrier d'apprentissage .	62
4.1	Représentation des croyances entrant en jeu dans la correction de l'information fournie par la source.	66
4.2	Illustration [27] d'un transfert de masse par marginalisation (en haut), et par extension vide (en bas), dans le cas d'un espace produit.	67
4.3	Illustration [27] d'un transfert de masse par conditionnement (en haut), et par déconditionnement (en bas), dans le cas d'un espace produit. . .	68
4.4	Illustration du transfert de masse par déconditionnement d'une masse définie sur Ω conditionnelle à R sur l'espace produit $\Omega \times \mathcal{R}$	70
4.5	Illustration d'un transfert de masse lors d'un affaiblissement contextuel.	78

4.6	Illustration des réductions extérieure et intérieure d'un sous-ensemble A de Ω avec un grossissement « en pavé » de Ω	80
4.7	Optimisation des coefficients d'affaiblissement lors d'une combinaison.	85
4.8	Illustrations des modèles conduisant aux mécanismes de correction introduit au paragraphe 4.7.	90
5.1	Éléments du grossissement employé pour réaliser un affaiblissement Θ -contextuel de la fonction de masse issue du LAP 3.	94
5.2	Illustration de l'amélioration au niveau distribution d'un point de fonctionnement obtenu grâce à un affaiblissement contextuel basé sur le grossissement illustré sur la figure 5.1.	95
5.3	Performances au niveau acheminement . Illustration d'un réglage R_0 plus intéressant que le réglage R_1 associé au point $CMCT_1$, et qui, grâce à l'affaiblissement contextuel, obtient les mêmes performances au niveau distribution que le point $CMCT_1$ (figure 5.2).	97
5.4	Scores et vérité au niveau acheminement des décisions manuscrites, différentes d'un rejet complet, fournies par les LAP 1 et 2 sur le lot de courrier d'apprentissage. Un losange foncé correspond à une décision erronée au niveau acheminement. Un carré clair est associé à une décision correcte au niveau acheminement.	98
5.5	Extension du modèle de base par la prise en compte de scores de confiance comme paramètres de correction de l'affectation postale employée.	99
5.6	Apprentissage d'une correction des fonctions de masse issues des LAP 1 et 2 par un affaiblissement classique avec un taux d'affaiblissement fonction du score attaché à la décision du LAP.	100
5.7	Introduction d'un mécanisme de correction des fonctions de masse issues des LAP 1 et 2.	102
5.8	Performances de la combinaison de base (losange foncé) et de la combinaison avec une correction par les scores des informations issues des LAP 1 et 2 (croix sur fond clair), réalisées sur l' ensemble d'apprentissage	103
5.9	Performances au niveau distribution de la combinaison de base (losange foncé) et de la combinaison avec une correction par les scores des informations issues des LAP 1 et 2 (croix sur fond clair), réalisées sur l' ensemble d'apprentissage	104
5.10	Performances de la combinaison de base (losange foncé) et de la combinaison avec une correction par les scores des informations issues des LAP 1 et 2 (croix sur fond clair), réalisées sur le lot de courrier de test	105
5.11	Scores et vérité au niveau distribution des décisions manuscrites fournies par le LAP 1 sur le lot de courrier d'apprentissage. Un losange foncé correspond à une décision erronée au niveau distribution. Un carré clair est associé à une décision correcte au niveau distribution.	106
5.12	Nombres de bonnes reconnaissances et d'erreurs en distribution manuscrite pour chaque réglage de la combinaison sans et avec prise en compte des scores selon le mécanisme (5.8), sur le lot de courrier d'apprentissage	107
5.13	Répartitions en acheminement dactylographié	108
5.14	Répartitions en acheminement manuscrit	108
5.15	Répartitions en distribution dactylographiée	109

5.16	Répartitions en distribution manuscrite	109
5.17	Performances des LAP, de la combinaison de base et des deux combinaisons prudentes obtenues sur le lot de courrier d'apprentissage	111
A.1	$C = (A \setminus B)^* \cup D$, où $D \in \mathcal{C}$, $D \subseteq B_*$, $C \in \mathcal{C}$ et $B \cup C = A$	133

Introduction

La reconnaissance automatique d'adresses postales est un problème difficile, recourant à une chaîne complexe de traitements variable suivant les pays [9]. Cette chaîne débute par l'acquisition d'une image d'un envoi sur lequel se trouve l'adresse du destinataire à identifier. Elle implique différents algorithmes de reconnaissance orientés vers le courrier manuscrit ou dactylographié. Elle se conclut par la sortie d'un code associé à une adresse postale se trouvant dans une base de données contenant l'ensemble des adresses postales du pays traité. Cette chaîne de traitement définit un *lecteur d'adresses postales (LAP)*, encore appelé *moteur OCR*, pour « Optical Character Recognition Engine ».

Une voie d'amélioration de la reconnaissance du courrier postal réside dans la combinaison ou fusion de différents LAP. En effet, une décision fondée sur un grand nombre d'informations d'origines et de natures variées est généralement plus robuste que toute décision prise individuellement par chaque source d'information [1, 13, 52, 114]. Le problème de fusion traité dans ce mémoire concerne l'agrégation d'informations dans le but d'améliorer la prise de décision [13, 79].

Une des caractéristiques importantes des informations à fusionner, justifiant l'intérêt d'une combinaison, concerne leurs imperfections. Les principales formes en sont l'*incertitude*, l'*imprécision* et l'*ambiguïté* [39, 55]. L'incertain est relatif à la vérité d'une proposition alors que l'imprécision est relative à la nature d'une proposition [33]. Par exemple, la proposition « Jean mesure entre 1,5 mètre et 2 mètres » est certaine et imprécise, alors que la proposition « Je crois que Jean mesure 1,5 mètre » est incertaine et précise. L'ambiguïté est une forme particulière d'imprécision, correspondant au passage graduel d'une catégorie à une autre. elle est due à une absence de rigueur ou à la flexibilité inhérente d'un système [14]. Par exemple, c'est le cas des propositions exprimées dans le langage courant : « Jean est grand », ou de propositions confrontées à des phénomènes naturels : le passage graduel du jour et de la nuit, la maturation d'un fruit. La théorie des ensembles flous, initiée par Zadeh [124], constitue un cadre bien adapté à la représentation et la manipulation de telles informations.

Les combinaisons actuelles d'adresses postales sont réalisées par l'intermédiaire de deux grands schémas : le *schéma de combinaison en cascade* et le *schéma de combinaison en parallèle*.

Dans le *schéma de combinaison en cascade*, illustré sur la figure 1(b), les LAP sont placés les uns à la suite des autres. L'image de l'envoi est transférée à un premier LAP. Si ce dernier a pu déterminer l'adresse de cet envoi, celle-ci est retournée, sinon l'image est transmise à un second LAP. Ce processus est itéré jusqu'à ce qu'un LAP ait pu déterminer une adresse ou que tous les LAP aient été utilisés. Si aucune

adresse n'a pu être déterminée, de nombreuses variantes existent pour déterminer une décision associée à une adresse incomplète fournie par un des LAP de la cascade. Lors de cette phase, le schéma de combinaison de ces variantes n'est plus un schéma en cascade, mais un schéma en parallèle.

Dans le *schéma de combinaison en parallèle*, illustré sur la figure 1(a), l'image de l'envoi est transférée simultanément à tous les LAP, ceux-ci transférant leurs décisions à un moteur de combinaison en charge de fournir une décision.

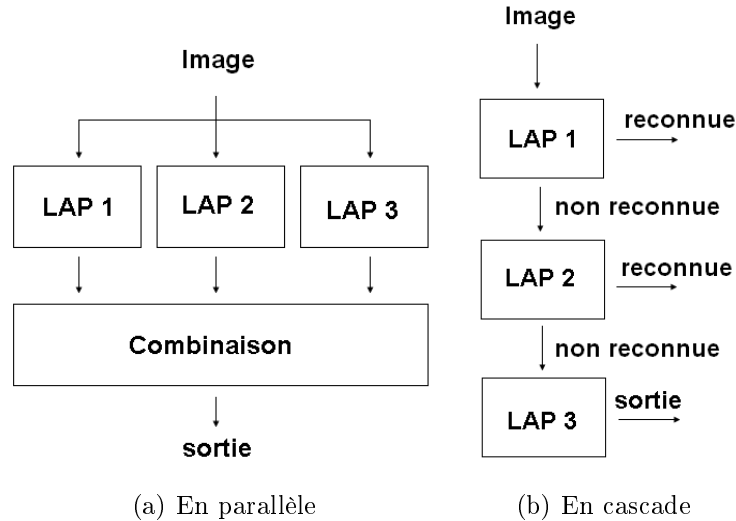


Figure 1 – Schémas de combinaison de trois lecteurs d'adresses postales.

À l'intérieur de ces deux grands schémas les combinaisons existantes abordent ce problème de fusion par l'intermédiaire de règles [18,42], consistant essentiellement à prendre en compte à la fois la redondance des informations fournies par les LAP ainsi que des informations statistiques relatives aux performances des LAP, afin de réaliser la meilleure combinaison possible. Par exemple, dans [42] l'idée consiste à choisir la décision majoritaire si elle existe, sinon la décision du LAP jugé le plus fiable est choisie. Bien que ces règles, issues du bon sens, peuvent se montrer performantes pour un nombre réduit de LAP, un travail approfondi sur une modélisation plus riche d'une fusion d'adresses postales est attendu.

Dans cette thèse, notre approche de cette combinaison postale s'est alors orientée vers le *modèle des croyances transférables* (MCT) développé par Smets [10]. Le MCT [103,111] est un modèle de représentation de connaissances partielles reposant sur une interprétation subjective de la théorie des fonctions de croyance de Dempster et Shafer [88,99]. Il permet de gérer des informations aussi bien incertaines qu'imprécises, et fournit nombre d'outils pour la modélisation d'une telle fusion [106].

Ce mémoire aborde donc la conception, la réalisation et les tests de modèles de fusion d'informations postales basés sur le modèle des croyances transférables. Il est composé de cinq chapitres.

Dans le premier chapitre, les notions et le vocabulaire relatifs au domaine de la reconnaissance des adresses postales sont abordés. Le problème de fusion d'adresses postales est décrit. Il se caractérise notamment par l'organisation hiérarchique des

décisions, et par la possibilité donnée aux lecteurs d'adresses postales de fournir une adresse postale complète ou partielle.

Le second chapitre est consacré à la présentation des principaux concepts du MCT employés dans nos travaux. L'idée principale de transfert des masses à des ensembles plus ou moins grossiers en fonction des informations disponibles est illustrée au travers des concepts de spécialisation et de généralisation. Un paragraphe de ce chapitre est dédié à la construction d'une fonction de croyance à partir d'une matrice de confusion.

Le troisième chapitre présente un modèle générique pour la fusion de décisions postales dans le cadre du MCT. Plusieurs méthodes d'affectation basées sur une matrice de confusion et une hiérarchie des décisions sont introduites. Le cadre de discernement choisi et la méthode de prise de décision sont détaillés. Les résultats obtenus par ce modèle montrent une amélioration significative par rapport aux performances individuelles des LAP, ainsi que par rapport aux résultats obtenus par des schémas à base de votes.

Dans le quatrième chapitre, de nouveaux outils de correction de fonctions de croyance sont développés et justifiés. Ceux-ci permettent de prendre en compte des informations relatives à la fiabilité d'une source d'information, avec plus de finesse que l'opération d'affaiblissement introduite par Shafer.

L'intérêt de ces mécanismes est illustré dans le cinquième chapitre, où sont présentées des améliorations du modèle de base exposé au chapitre 3. Celles-ci exploitent de nouvelles informations complémentaires des décisions fournies par les LAP, notamment, des informations de fiabilité contextuelle et des scores de confiances fournies par certains LAP. Des résultats probants obtenus par ces modèles sont décrits. Enfin, une perspective à court terme permettant d'entrevoir le potentiel de cette combinaison, vient clore ce chapitre.

Une synthèse des apports de cette thèse est réalisée dans la conclusion de ce mémoire. De nombreuses perspectives sont évoquées pour améliorer encore cette fusion postale.

Ce travail est le fruit d'une collaboration entre le laboratoire Heudiasyc de l'Université de Technologie de Compiègne et l'entreprise Solystic, réalisée dans le cadre de la convention CIFRE 466/2003. Il fait l'objet d'une demande de brevet déposée à l'INPI le 17 mars 2006 [63].

Introduction au domaine postal

Dans ce chapitre, les notions et le vocabulaire relatifs au domaine de la reconnaissance des adresses postales sont abordés. La problématique de fusion de décisions de lecteurs d'adresses postales (LAP) est exposée.

1.1 Concept d'adresse postale et définitions

1.1.1 Structure générale d'une adresse postale

Définition 1.1 (Adresse postale). *Une adresse postale est une chaîne de caractères permettant à un destinataire de recevoir un envoi postal (lettre, courrier, colis) d'un expéditeur via une organisation postale attachée à un pays.*

Chaque pays a structuré ses adresses postales d'une manière spécifique, néanmoins tous ces prototypes d'adresses possèdent une structure *hiérarchisée*. De bas en haut, chaque ligne précise le lieu de destination.

Une adresse postale est composée de deux grandes parties :

- la *partie acheminement*, qui indique le bureau de distribution, en général la ville de destination de l'envoi postal ;
- la *partie distribution*, qui indique le lieu précis de délivrance de l'envoi.

La partie distribution est relative :

- soit à un pas de porte dans une voie (rue, avenue, boulevard, place) ; dans ce cas nous parlerons de *distribution géographique* (Tableau 1.1) ;

Tableau 1.1 – Exemples d'adresses françaises de même acheminement, mais de distributions géographiques différentes.

Jean D. 7 Rue de l'Oise 60200 COMPIEGNE FRANCE	Agence D 26 Place du change 60200 COMPIEGNE
---	---

- soit à des mentions spéciales de distribution dépendantes du pays traité ; dans ce cas nous parlerons de *distribution spéciale*. Par exemple, en France, nous trouvons parmi les distributions spéciales : les boîtes postales, TSA (Tri Service Arrivée), Autorisation (Tableau 1.2). Ces intitulés marquant une distribution spéciale sont appelés *séparateurs*.

Tableau 1.2 – Exemple d’adresses françaises comportant des distributions spéciales.

Entreprise Dupont & Dupond B.P. 11 84004 AVIGNON Cedex	Société X AUTORISATION 39842 75670 PARIS Cedex
Transport Y BOITE POSTALE 63 75662 PARIS Cedex 14	Assurance Z TSA 35071 59714 LILLE Cedex 9

Dans chaque pays, une adresse postale est identifiée par un code, appelé *code d’indexation*, issu d’une base de données contenant l’ensemble des adresses postales et des codes d’indexation du pays concerné.

Un code d’indexation est composé d’un code acheminement et d’un code distribution. Ce code varie suivant les pays.

Exemple 1.1. *Au Royaume-Uni, le code acheminement est variable, il est composé de deux, trois ou quatre caractères alphanumériques. Le code distribution est composé de trois caractères. En Hollande, le code acheminement est composé de quatre chiffres et le code distribution de deux lettres. Au Brésil, le code acheminement est composé de quatre chiffres et le code distribution de trois chiffres. Dans ces pays, le code d’indexation est indiqué dans l’adresse (Tableau 1.3). En France, le code acheminement est composé de cinq chiffres généralement égaux au code postal de la ville du destinataire. Le code distribution est composé de cinq caractères, et contrairement au code acheminement, il n’est jamais mentionné dans l’adresse.*

Tableau 1.3 – Exemples d’adresses traitées respectivement au Royaume-Uni, en Hollande et au Brésil.

James B. 2 Rougier Street YORK Y090 2UU	Nicolas H. Oranje Nassaulaan 13 6026 BW MAARHEEZE	Edson Arantes D.N. R. Antonio de mariz, 132 05060-010 SAO PAULO - SP
--	---	--

Dans ce mémoire, les codes d’indexation seront simplifiés. Nous noterons :

- une partie acheminement par un code A_i ,
- une boîte postale ou tout autre distribution spéciale par un code B_j ,
- une rue ou tout autre voie géographique par R_k ,

où i , j et k sont des indices. Le numéro de pas de porte se notera à la suite du code de la rue.

Exemple 1.2. *Le code d’indexation « A_2R_26 » est associée à une adresse comportant une distribution géographique : le numéro 6 rue R_2 dans A_2 . Par exemple, il pourrait s’agir de l’adresse « 6 avenue Victor Ségoffin - 31400 Toulouse ».*

De même, « A_3B_1 » est le code d'indexation d'une adresse comportant une distribution spéciale : la boîte postale B_1 dans A_3 . Par exemple, il pourrait s'agir de l'adresse « Assurance Z - TSA 35071 - 59714 LILLE Cedex 9 ».

Dans ce mémoire, le pays choisi comme cadre applicatif de nos travaux est la France. Ainsi, la structure des adresses françaises est détaillée dans le paragraphe suivant.

1.1.2 Structure détaillée des adresses postales françaises

1.1.2.1 Acheminement non distribuable

En France, certains codes d'indexation sont uniquement composés d'un code acheminement. Ces lieux d'acheminement correspondent en général :

- à des petites villes ou villages ; par exemple, le code d'indexation de l'adresse « 2 le Bourg - 33230 Saint Christophe de Double » est composé uniquement d'un code acheminement ;
- à certains codes postaux cedexés ; par exemple, le code d'indexation de l'adresse « Martin & co. - BP 11 - 60201 Compiègne Cedex » est de même composé uniquement d'un code acheminement.

Ces lieux d'acheminement seront dit *non distribuables* : la partie distribution n'est pas à préciser pour obtenir le code d'indexation de l'adresse. Le caractère distribuable d'un acheminement est indiqué par la base de données postales.

1.1.2.2 Voie non numérotée

Dans certaines voies, tous les pas de porte sont codés par le même code distribution. Ainsi, il est inutile de préciser le numéro de pas de porte pour obtenir le code distribution. Seules la ville et la rue doivent être détectées pour obtenir le code d'indexation d'une adresse comportant ce type de distribution.

Exemple 1.3. Une sortie de LAP en faveur du code « A_3R_1 » codera tous les pas de porte de la rue R_1 dans A_3 . Que l'adresse indique le numéro 7 ou 9 ou 17 dans cette rue, le code d'indexation de cet envoi sera « A_3R_1 ». La nécessité de préciser le numéro de voie est indiquée par la base de données.

Ces voies, dont le numéro n'est pas à préciser, seront dites *non numérotées*. Nous parlerons de distribution géographique voie non numérotée, et par opposition, de distribution géographique voie numérotée, lorsque le numéro de pas de porte devra être précisé.

Remarque 1.1. L'ensemble des adresses postales et l'ensemble des codes d'indexation ne sont par conséquent pas en correspondance biunivoque : plusieurs adresses postales, au sens de la définition 1.1, peuvent être associées à un même code d'indexation.

En résumé, les adresses postales françaises peuvent être regroupées dans les catégories suivantes :

- les adresses comportant un acheminement non distribuable, associées à un code d'indexation de la forme « A_i » ;
- les adresses comportant une distribution géographique voie non numérotée, associées à un code d'indexation de la forme « $A_i R_k$ » ;
- les adresses comportant une distribution géographique voie numérotée, associées à un code d'indexation de la forme « $A_i R_k N$ », où N est un nombre ;
- les adresses comportant une distribution spéciale, associées à un code d'indexation de la forme « $A_i B_j$ ».

1.2 Lecteurs d'adresses postales (LAP)

Un *lecteur d'adresses postales (LAP)* est défini comme un système, associé à une base de données, en charge de la reconnaissance d'une adresse postale à partir d'une image d'un envoi postal.

1.2.1 Des décisions multi-niveaux

La reconnaissance complète d'une adresse postale étant une tâche complexe, un LAP est autorisé à fournir des adresses postales partielles. Par exemple, lorsqu'une adresse postale comporte une distribution géographique voie numérotée, le LAP peut :

- s'arrêter au niveau de la rue en ne précisant pas le numéro de pas de porte, dans ce cas nous dirons que le numéro de pas de porte a été *rejeté* ;
- s'arrêter au niveau de la partie acheminement, en ne précisant ni la rue, ni le numéro de porte, dans ce cas nous dirons que la rue a été *rejetée* ;
- ne préciser aucun élément de l'adresse, et dans ce cas nous dirons que la ville a été *rejetée*. Cette dernière décision est un « rejet complet ».

Ces adresses partielles sont retranscrites par un code d'indexation, où la partie rejetée est notée par « *Rej* ».

Exemple 1.4. *Supposons qu'un LAP ait bien reconnu une adresse en faveur du numéro 6 de la voie numérotée R_2 dans A_2 . Alors il peut fournir le code d'indexation attachée à cette adresse : « $A_2 R_2 6$ ».*

Supposons que le LAP n'ait pas pu identifier le numéro de voie, alors il pourra décider une adresse partielle comportant uniquement la ville et la rue, et ainsi fournir le code d'indexation « $A_2 R_2 Rej$ ».

De même, s'il n'a reconnu que la ville, il pourra décider « $A_2 Rej$ ».

*Enfin, si la ville n'a pu être déterminée, il décidera un rejet complet retranscrit par le code « *Rej* ».*

L'ensemble des décisions, c'est-à-dire des codes d'indexation, qu'un LAP peut fournir est ainsi organisé de manière hiérarchique. Les niveaux sont les suivants :

- le niveau rejet complet associée à la décision la plus imprécise ;
- le niveau acheminement, où se trouvent toutes les décisions en faveur d'acheminements non distribuables : ce sont des adresses complètes, et toutes les décisions en faveur d'adresses partielles où seule la ville est précisée ;

- le niveau rue, où se trouvent, par convention, uniquement les décisions en faveur de voie numérotée où seule la ville et la rue sont précisée ;
- le niveau distribution, où se trouvent toutes les décisions en faveur des adresses complètes de catégories distributions spéciales et distributions géographiques.

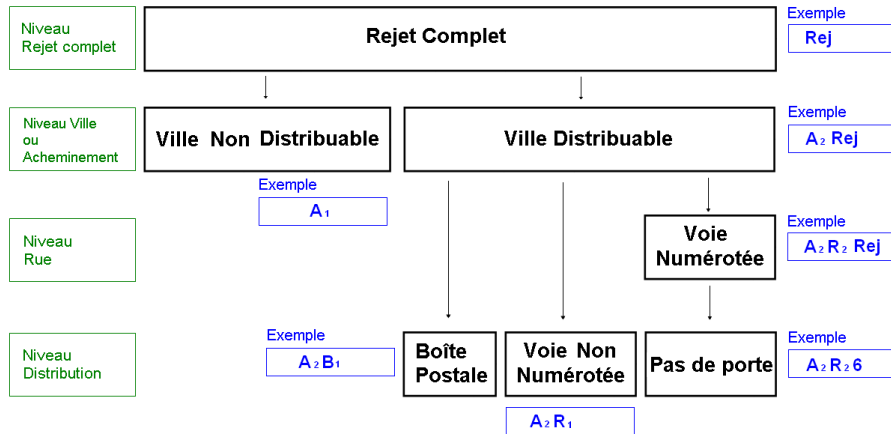


Figure 1.1 – Niveaux de décisions autorisées pour chaque LAP avec des exemples associés.

La figure 1.1 illustre la hiérarchie des niveaux de décisions autorisées pour chaque LAP. À chaque niveau sont représentées des catégories d'adresses avec des exemples de décisions associées.

Il serait possible de modifier cette hiérarchie postale, en enlevant ou ajoutant des niveaux et des catégories de décisions, par exemple :

- ajouter un niveau pays (un niveau intermédiaire entre le rejet complet et l'acheminement), en distinguant les pays étrangers ;
- ajouter des zones intermédiaires entre le pays et l'acheminement (niveau département, région, état, ...);
- distinguer les villes étrangères des villes du pays traité ;
- distinguer certains groupes de séparateurs dans les cas de distribution spéciale.

Dans ce mémoire, la hiérarchie représentée sur la figure 1.1 fait office de hiérarchie de référence afin d'illustrer nos travaux. En pratique, cette hiérarchie dépend de la structure des adresses postales du pays client dans lequel est trié le courrier, des LAP à notre disposition, et de la représentativité des catégories.

Un rôle important dans la reconnaissance d'une adresse postale est joué par le type d'écriture des courriers.

1.2.2 Types d'écriture d'un courrier

L'ensemble du courrier peut se diviser en deux catégories : le courrier *dactylographié* et le courrier *manuscrit*.

Les lettres dactylographiées constituent la source majeure du courrier : environ 75 % du courrier à traiter dans les pays européens est dactylographié. L'uniformité de l'écriture (espacement des caractères, localisation des mots) définit la caractéristique

principale de ce type de courrier. Une reconnaissance individuelle des mots est possible.

À l’opposé, la reconnaissance d’une adresse manuscrite, où une infinité de « polices » existent, est beaucoup plus complexe. Les caractères sont plus ou moins attachés. Les espacements entre les mots sont plus ou moins réguliers. La segmentation en caractères et la reconnaissance des caractères est ainsi plus difficile. Une reconnaissance individuelle de chaque mot n’est pas toujours possible.

La figure 1.2 illustre des exemples de courriers dactylographiés et manuscrits.

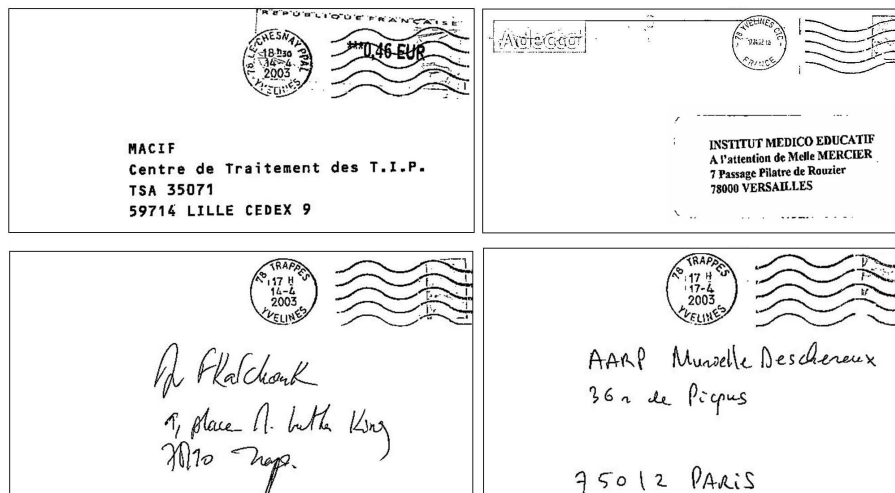


Figure 1.2 – Exemples de courriers dactylographiés et manuscrits.

Les LAP Solystic emploient des algorithmes différents pour la reconnaissance du courrier dactylographié et manuscrit. Cette stratégie est certainement de même adoptée par les LAP de la concurrence. Notons enfin qu’il existe des LAP spécialisés dans la reconnaissance d’un seul type d’écriture.

1.2.3 Modélisation générique d’un LAP

Les LAP à fusionner sont issues de Solystic et de la concurrence. Les LAP Solystic peuvent fournir en plus de leur décision d’autres indications. Par exemple, en plus du type d’écriture du courrier traité, il est possible d’obtenir des informations de corrélations entre mots, des scores de reconnaissance de caractères ou de mots issues des différents algorithmes employés. Cependant, les LAP d’origine extérieure ne fourniront qu’une décision avec le type d’écriture du courrier traité pour seule indication.

Différentes stratégies de fusion peuvent être employées. Il est possible d’intégrer toutes les informations issues des LAP Solystic pour construire avec les codes d’indexation issus des LAP extérieurs disponibles une nouvelle décision. Or cette construction est complexe, et il n’est pas souhaité de construire à ce jour un nouveau LAP, dont la mise en œuvre nécessite plusieurs années. Ainsi, le travail de fusion va être effectué par étape.

Dans un premier temps, le problème posé concerne la construction d'une combinaison générique. Les LAP Solystic et extérieurs sont considérées comme des boîtes noires (figure 1.3) liées à une base de données d'adresses postales, prenant en entrée une image d'un envoi postal sur laquelle se trouve l'adresse postale à identifier, et retournant en sortie :

- une décision en faveur d'un code d'indexation, issu de la base de données postales, associé à une adresse postale partielle ou complète ;
- une indication sur le type d'écriture du courrier traité.

Cette combinaison pourra en particulier ne combiner que des LAP extérieurs, et ainsi constituer un nouveau produit Solystic pouvant être employé indépendamment des LAP Solystic.

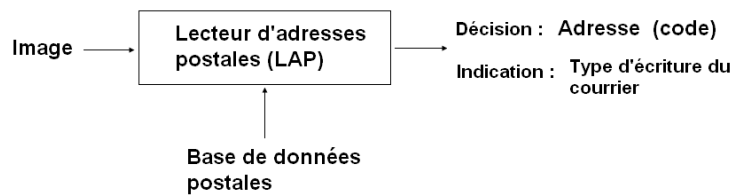


Figure 1.3 – Modélisation d'un lecteur d'adresses postales.

Finalement, le problème traité est un problème de fusion de décisions postales.

1.3 Problématique de la fusion des décisions postales

Supposons que nous disposions de N LAP LAP_i , $i \in \{1, \dots, N\}$, utilisant la même base de données postales et associés à la même hiérarchie. La problématique de fusion consiste à fournir, pour chaque image, la « meilleure décision » possible connaissant les N sorties des LAP, c'est-à-dire, les N décisions et les N indications des LAP.

Exemple 1.5. *Considérons 3 LAP, notés LAP_1 , LAP_2 , et LAP_3 , ayant formulé les décisions d_1 , d_2 , et d_3 suivantes au regard d'une image d'un envoi postal :*

- $d_1 = \text{« } A_2\text{Rej »}$, courrier dactylographié ;
- $d_2 = \text{« } A_1 \text{ »}$, courrier dactylographié ;
- $d_3 = \text{« } A_2R_{17} \text{ »}$, courrier dactylographié.

Notre problème de fusion postale consiste alors à répondre à la question suivante : quelle décision associée à quel niveau devons nous prendre ? Une décision de niveau acheminement majoritaire : « $A_2\text{Rej}$ » ? Une adresse complète « A_1 » ou « A_2R_{17} » ? La décision la plus prudente « Rej » ? ou bien, de ces trois propositions pourrait émerger une nouvelle décision, par exemple « A_3 » ?

Pour statuer sur cette meilleure décision, des critères de performance sont définis de la manière suivante.

Définition 1.2 (Taux de performance). À chaque niveau de la hiérarchie (par exemple au niveau acheminement ou au niveau distribution), il est possible de mesurer les performances d'un LAP ou de la combinaison en termes de taux de lecture, d'erreur et de rejet :

- le taux de lecture ou taux de reconnaissance est égal au pourcentage de bonnes décisions données par le LAP à ce niveau ;
- le taux d'erreur ou taux de substitution est égal au pourcentage de mauvaises décisions données par le LAP à ce niveau ;
- le taux de rejet est égal au pourcentage de rejets effectués par le LAP à ce niveau.

Ainsi, à chaque niveau, la somme de ces trois taux est égale à 1.

Généralement les LAP sont comparés selon leurs taux de performance (taux de lecture et d'erreur) en acheminement et en distribution sachant que :

- les performances en distribution concernent des images dont la vérité est relative à une rue ou une boîte postale, les villes non distribuables ne sont pas concernées ;
- les performances en acheminement concernent toutes les images. Dès que le point acheminement est correct, la décision est considérée comme correcte au niveau acheminement même si la distribution est erronée. Par exemple, une décision « A_2R_1 » fournie sur une image dont la vérité est « A_2R_21 » est correcte au niveau acheminement, en effet, la lettre sera acheminée dans la bonne ville. Mais, cette décision est en erreur au niveau distribution : la lettre sera distribuée dans la mauvaise rue.

La lecture du courrier jusqu'au point distribution est ainsi plus difficile que la seule lecture de la partie acheminement. Le tableau 1.4 donne une idée des performances moyennes des systèmes Solystic appliqués à la France. Notons que ces performances dépendent de la qualité du courrier. Ces performances sont obtenues sur des lots comportant plus de 50% de « bon » courrier, issu de sociétés remplissant des normes de qualité de rédaction en échange de réduction de tarifs postaux.

Tableau 1.4 – Performances moyennes réalisées par les LAP Solystic en fonction du type de courrier et du niveau.

	Courrier dactylographié		Courrier manuscrit	
	Taux de lecture	Taux d'erreur	Taux de lecture	Taux d'erreur
Acheminement	$\geq 95\%$	$\leq 0.5\%$	$\geq 80\%$	$\leq 1\%$
Distribution	$\geq 80\%$	$\leq 1\%$	$\geq 40\%$	$\leq 2\%$

Les performances de chaque LAP peuvent être représentées sur un graphique avec en abscisses le taux de lecture, et en ordonnées le taux d'erreur. Ce graphique permet de visualiser et de déterminer d'une manière simple quel LAP présente les meilleures performances et quels LAP ne sont pas comparables.

Exemple 1.6. La figure 1.4 représente les performances de quatre LAP. Le LAP 4 possède les meilleures performances : il lit plus que les autres et commet le moins

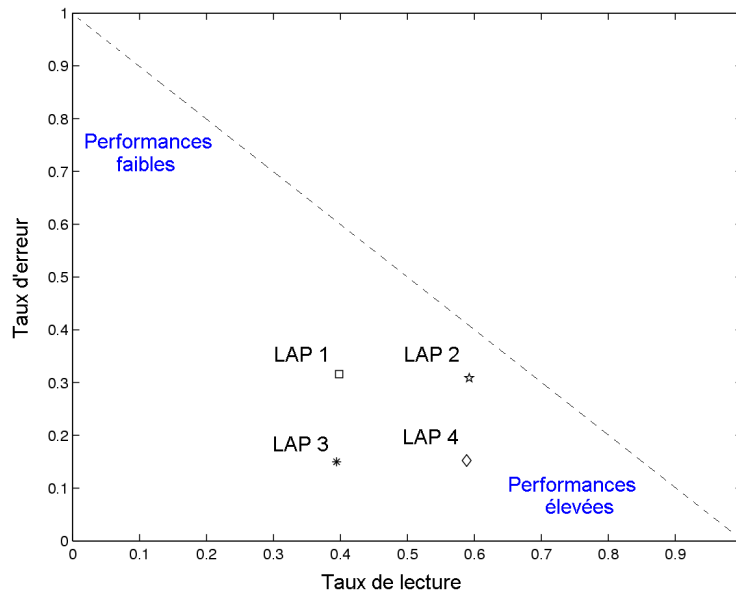


Figure 1.4 – Représentation graphique des performances à un certain niveau de quatre LAP.

d'erreurs. Les LAP 2 et 3 possèdent des performances non comparables, le premier lit plus que le second mais il commet un plus grand nombre d'erreurs.

L'objectif des modèles de combinaison réside dans l'obtention de performances se situant dans la zone élevée par rapport aux performances des LAP à fusionner.

1.4 Conclusion

Dans ce chapitre, notre problème de fusion a été défini. Il s'agit d'un problème de fusion de décisions postales, dont l'originalité réside dans l'organisation hiérarchique des décisions, chaque LAP pouvant fournir une décision d'un niveau quelconque dans la hiérarchie. La combinaison devra trouver un compromis entre, fournir une adresse complète, qui sera excellente seulement si elle est correcte, et, fournir une adresse partielle apportant moins d'informations qu'une adresse complète, mais dont on sera toujours plus sûr qu'elle soit correcte.

Le chapitre suivant présente les principaux concepts du modèle des croyances transférables choisi comme cadre théorique pour la réalisation de cette fusion.

Modèle des croyances transférables

2.1 Introduction

Même si la notion de non-additivité des degrés de croyance peut se rencontrer dans la littérature du XVIII^e siècle dans le traitement de probabilités de témoignages (travaux de Hooper, 1699, *A calculation of the Credibility of Human Testimony*, de J. Bernoulli, 1713, *Ars Conjectandi*, et de J.H. Lambert, 1764, *Neues Organon*) [11, 89–92], la théorie moderne des fonctions de croyance trouve ses origines à la fin des années 60 dans les travaux de Dempster [20–22] sur la généralisation de l'inférence Bayésienne lorsqu'il n'y a pas d'a priori sur les paramètres. Cette généralisation conduit à l'utilisation de probabilités non additives pouvant être combinées par une règle, qui portera plus tard le nom de règle de Dempster. Dans son livre référence de 1976, *A Mathematical Theory of Evidence* [88], Shafer présente les fonctions de croyance comme un cadre général de représentation des incertitudes, englobant la théorie des probabilités comme cas particulier. Pour cela, il s'appuie sur deux idées principales : un degré de croyance entre 0 et 1, et une combinaison des croyances par la règle de Dempster. Ce travail de synthèse et de simplification conduit à faire sortir cette théorie hors du cadre restreint de l'inférence statistique. Dès le début des années 80, cette théorie, devenue « théorie de Dempster-Shafer », fait son entrée dans le domaine de l'intelligence artificielle [6, 44], où le besoin de modélisation de raisonnements de plus en plus complexes a montré les limites des modèles probabilistes en présence d'informations pauvres ou arbitraires. Ce domaine continue de motiver l'expansion de cette théorie [104]. Sans être exhaustif, mais afin d'illustrer la diversité des applications et l'actualité de cette théorie, notons que les fonctions de croyance ont déjà été employées dans les systèmes experts [4, 93], en reconnaissance des formes [23, 25, 26], en analyse de risque : évaluation du risque de tremblement de terre [43], de la qualité des eaux usées en présence d'informations pauvres [30], en fusion multi-capteurs [2, 7, 41], en fusion de classifieurs [62, 77, 120], en imagerie médicale [11], satellitaire ou radar [59, 80, 115], en détection et suivi de cible [3, 5, 75, 76], en robotique [73] et véhicules intelligents [45, 81]. De nombreuses autres références d'applications ont été recensées par Sentz et Ferson dans [87] en 2002, et un état de l'art sur les fonctions de croyance a été réalisé par Vannoorenberghe en 2003 [117].

Dès 1978 [94], Smets a axiomatisé et clarifié les bases inscrites dans le livre de Shafer, et a développé le modèle des croyances transférables (MCT) [100, 103, 111]. Ce modèle constitue une interprétation subjective et non probabiliste de la théorie de Dempster-Shafer, dans laquelle une fonction de croyance représente une opinion

pondérée d'un agent, aucune mesure de probabilité n'étant sous-entendue dans la représentation des informations. D'autres modèles utilisent le concept mathématique de fonction de croyance, comme le modèle des Hints de Kolhas et Monney [56] et le modèle PAS de Haenni, Kohlas et Lehmann [47], basés sur le modèle de Dempster [20, 21]. Ces modèles constituent une autre interprétation des travaux de Dempster et Shafer [99]. Enfin, d'autres théories non-additives continuent de se développer comme par exemple la théorie des probabilités imprécises [118, 126].

Dans le cadre du MCT, deux niveaux sont distingués dans la modélisation du raisonnement de l'agent rationnel en charge de la prise de décision [39] :

- le niveau *crédal*, où sont représentées et manipulées les informations disponibles ;
- le niveau *pignistique* ou *décisionnel*, siège de la construction de la décision de l'agent.

Les intitulés de ces niveaux font référence à des mots latins, « credo », je crois, et « pignus », un pari [111, 112].

2.2 Représentation des connaissances

2.2.1 Concepts de base

La connaissance d'un agent est modélisée par l'allocation d'une masse finie de croyance à des sous-ensembles de l'univers de discours. Lorsque toutes ces masses de croyance portent sur des singletons, cette représentation est une distribution de probabilité. Mais en général, par manque d'informations, ces masses de croyance ne peuvent pas être affectées à des sous-ensembles aussi restreints.

Formellement, étant donnée x une variable à valeurs dans un ensemble fini¹ $\Omega = \cup_{k=1}^K \{\omega_k\}$, appelé *cadre de discernement* ou *univers*. La connaissance détenue par un agent rationnel Ag , quant à la valeur ω_0 effectivement prise par x au temps t , étant donné un ensemble de connaissances ou d'informations, appelé corpus d'évidences et noté EC , peut être quantifiée par une *fonction de masse* définie sur l'ensemble des parties de Ω , noté 2^Ω , à valeurs dans $[0, 1]$, et vérifiant la condition suivante :

$$\sum_{A \subseteq \Omega} m_{Ag,t}^\Omega\{x\}[EC](A) = 1 . \quad (2.1)$$

Remarque 2.1. *En l'absence d'ambiguïté, la notation complète sera simplifiée en m_{Ag}^Ω , m_{Ag} , m^Ω , ou même m . Par abus, nous dirons parfois que m^Ω est définie sur Ω , au lieu de sur l'ensemble des parties de Ω .*

Interprétation : une masse $m_{Ag}^\Omega(A)$ représente la *part de croyance* de l'agent Ag en la proposition « $\omega_0 \in A$ ».

Initialement, Shafer instaura une condition supplémentaire dans la définition d'une fonction de masse : la masse sur le vide doit être nulle [88, page 38]. Cette

¹Remarquons que le MCT a aussi été étendu à des cadres de discernement infinis [108].

condition n'existe pas dans le MCT. L'interprétation de $m(\emptyset)$ est discutée dans ce chapitre au paragraphe 2.4.9.

Une fonction de masse m est en correspondance biunivoque avec les fonctions suivantes.

$$\begin{aligned}
(\text{fonction de croyance}) \quad & bel(A) = \sum_{B:\emptyset \neq B \subseteq A} m(B) , \\
(\text{fonction de plausibilité}) \quad & pl(A) = \sum_{B:B \cap A \neq \emptyset} m(B) = bel(\Omega) - bel(\overline{A}) , \\
(\text{fonction de communalité}) \quad & q(A) = \sum_{B:B \supseteq A} m(B) , \\
(\text{fonction d'implicabilité}) \quad & b(A) = \sum_{B:B \subseteq A} m(B) = bel(A) + m(\emptyset) .
\end{aligned} \tag{2.2}$$

Le nombre $bel(A)$ représente la part totale spécifique de croyance soutenant une proposition A . Comme la masse allouée à l'ensemble vide soutient A et son complémentaire \overline{A} dans Ω , la masse sur le vide n'est comptée ni pour $bel(A)$, ni pour $bel(\overline{A})$, d'où la qualification de part totale de croyance spécifique. Remarquons que $bel(\Omega) < 1$ est possible dans le MCT.

La quantité $pl(A)$ représente la part maximale de croyance qui pourrait soutenir la proposition A si de nouvelles informations étaient disponibles.

Par rapport au pivot que représente la fonction de masse, les fonctions q et b ont une interprétation très symétrique. $b(A)$ représente la somme des masses allouées aux sous-ensembles de A , et $q(A)$ celle des masses allouées aux sur-ensembles de A . Ces fonctions jouent un rôle mathématique important dans l'usage des fonctions de croyance.

Toutes ces fonctions représentent la même information sous une forme différente. Ainsi, par abus de langage, le terme de fonction de croyance est employé pour désigner l'ensemble de ces fonctions.

2.2.2 Définitions

Définition 2.1 (Éléments focaux). *Un sous-ensemble A de Ω tel que $m(A) > 0$ est appelé élément focal de m .*

Définition 2.2 (Fonction de masse catégorique). *Une fonction de masse catégorique sur un sous-ensemble B de Ω est définie par :*

$$m_B(A) = \begin{cases} 1 & \text{si } A = B , \\ 0 & \text{sinon.} \end{cases} \tag{2.3}$$

Remarque 2.2 (Notations). *La notation m_B désigne une fonction de masse catégorique sur B si et seulement si B est un sous-ensemble de Ω . Sinon, m_{Ag} désigne une fonction de masse issue de l'agent Ag . Dans ce mémoire, l'origine des fonctions de croyance est, par exemple, un agent rationnel Ag , une source d'information S , un LAP.*

Définition 2.3 (Fonction de masse vide). *La fonction de masse vide est la fonction de masse catégorique sur Ω . Elle représente l'ignorance totale : $m_\Omega(\Omega) = 1$.*

Définition 2.4 (Fonction de masse Bayésienne). *Une fonction de masse est dite Bayésienne si et seulement si tous ses éléments focaux sont des singletons de Ω .*

Définition 2.5 (Fonction de masse consonante). *Une fonction de masse est dite consonante si et seulement si tous ses éléments focaux sont emboîtés.*

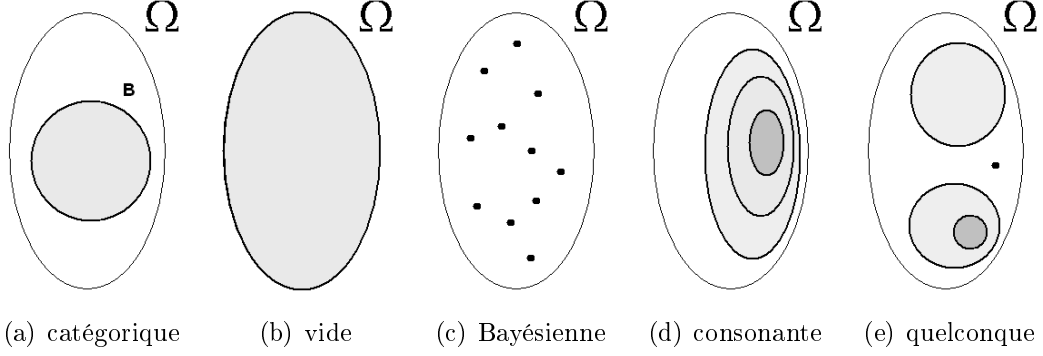


Figure 2.1 – Représentation des éléments focaux de certaines classes de fonctions de croyance.

La figure 2.1 illustre la forme des éléments focaux de ces classes de fonctions de croyance.

Définition 2.6 (Fonction de masse normalisée). *Une fonction de masse m est dite normalisée si et seulement si $m(\emptyset) = 0$.*

Définition 2.7 (Fonction de masse dogmatique). *Une fonction de masse m est dite dogmatique si et seulement si $m(\Omega) = 0$.*

2.2.3 Représentation matricielle

Une fonction de croyance définie sur 2^Ω peut être représentée par un vecteur colonne² de dimension $2^{|\Omega|}$, telle que ses éléments soient rangés dans un ordre binaire. En notant \mathbf{m} le vecteur correspondant à une fonction de masse m , le tableau 2.1 illustre un exemple de représentation vectorielle d'une fonction de masse définie sur un cadre de discernement $\Omega = \{a, b, c\}$. De même, les vecteurs \mathbf{b} , \mathbf{q} , \mathbf{bel} et \mathbf{pl} denoteront respectivement les vecteurs des fonctions d'implicabilité, de communalité, de croyance et de plausibilité. Enfin, l'ensemble des fonctions de masse définies sur Ω sera noté \mathcal{M}^Ω .

Les transformations entre les différentes représentations d'une fonction de croyance peuvent alors être modélisées par des matrices carrées d'ordre $2^{|\Omega|}$. Par exemple, la relation classique $b(A) = \sum_{B \subseteq A} m(B)$ peut s'écrire :

$$b(A) = \sum_{B \subseteq \Omega} BfrM(A, B) m(B) \quad (2.4)$$

²Tous les vecteurs utilisés dans ce mémoire seront par défaut des vecteurs arrangés en colonne, comme c'est usuellement le cas en algèbre linéaire.

Tableau 2.1 – Ordre des éléments d'un vecteur de masses associé à une fonction de masse définie sur $\Omega = \{a, b, c\}$.

<i>Position</i>	<i>cba</i>	<i>m</i>
1	000	$m(\emptyset)$
2	001	$m(\{a\})$
3	010	$m(\{b\})$
4	011	$m(\{a, b\})$
5	100	$m(\{c\})$
6	101	$m(\{a, c\})$
7	110	$m(\{b, c\})$
8	111	$m(\{a, b, c\})$

où, $\forall A, B \subseteq \Omega$,

$$BfrM(A, B) = \begin{cases} 1 & \text{si } B \subseteq A, \\ 0 & \text{sinon.} \end{cases} \quad (2.5)$$

En posant $\mathbf{BfrM} = [BfrM(A, B)]$, $A, B \subseteq \Omega$, on a $\mathbf{b} = \mathbf{BfrM} \cdot \mathbf{m}$.

Exemple 2.1. Illustration de la transformation d'une fonction de masse en une fonction d'implicabilité, en utilisant la matrice \mathbf{BfrM} avec $\Omega = \{a, b, c\}$. Les accolades des sous-ensembles de Ω ont été omises, par exemple, $\{a, b\}$ est noté a, b . L'absence de valeur dans la matrice indique un 0.

$$\mathbf{b} = \mathbf{BfrM} \cdot \mathbf{m}$$

	\emptyset	a	b	a, b	c	a, c	b, c	a, b, c	
\emptyset	=	$\begin{pmatrix} .0 \\ .0 \\ .1 \\ .1 \\ .3 \\ .3 \\ .6 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & & & & & \\ 1 & 1 & & & & & & & \\ 1 & & 1 & & & & & & \\ 1 & 1 & 1 & 1 & & & & & \\ 1 & & & & 1 & & & & \\ 1 & 1 & & & 1 & 1 & & & \\ 1 & & 1 & & 1 & & 1 & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \end{pmatrix}$	$\begin{pmatrix} .0 \\ .0 \\ .1 \\ .0 \\ .3 \\ .0 \\ .2 \\ .4 \end{pmatrix}$	\emptyset				
a					a				
b					b				
a, b					a, b				
c					c				
a, c					a, c				
b, c					b, c				
a, b, c					a, b, c				

La notation \mathbf{BfrM} , comme « \mathbf{b} from \mathbf{m} », identifie la matrice de transformation d'une fonction de masse en sa fonction d'implicabilité correspondante. Les autres transformations, pouvant aussi être représentées par des matrices de transformation, seront notées d'une façon similaire. Par exemple, \mathbf{QfrM} désigne la matrice de transformation de la fonction de masse en sa fonction de communalité correspondante. Les expressions de toutes ces transformations ont été regroupées par Smets dans [107].

2.3 Affectation d'une fonction de croyance par matrice de confusion

En pratique, construire une fonction de croyance représentant les informations disponibles n'est pas chose aisée, et il ne semble pas exister de solution universelle.

Généralement, cette phase de génération est très dépendante de l'application visée et de la nature des sources employées. Ainsi, dans de nombreuses applications en rapport direct avec un problème réel, les classifieurs sont jugés suffisamment connus pour que les éléments focaux soient fixés et les masses déterminées à partir de seuils ou d'abaques caractéristiques du classifieur employé [12, 49, 72, 80, 116]. En reconnaissance des formes, des méthodes d'affectation ont été définies à partir du vecteur forme de l'objet à identifier, par le calcul d'une distance [23, 26] ou d'une vraisemblance [2, 3, 119] [60, chapitre 3]. Notons aussi l'existence de méthodes d'affectations basées sur l'élicitation d'avis d'experts [8].

Dans le cadre de notre application, les classifieurs ne fournissent que des décisions brutes, et des données d'apprentissage sont disponibles. Il est alors possible de convertir la décision d'un classifieur en une fonction de croyance à l'aide de sa matrice de confusion, qui regroupe toutes ses décisions passées au regard d'objets dont la vérité est connue [120].

Dans ce paragraphe, des méthodes génériques de construction de fonctions de croyance à partir de matrices de confusion sont présentées.

2.3.1 Matrice de confusion

Une *matrice de confusion* $\mathcal{M} = (n_{k\ell})_{k \in \{1, \dots, K+1\}, \ell \in \{1, \dots, K\}}$ associée à un classifieur est une table décrivant les performances de ce classifieur sur un ensemble de test. Chaque ligne k , correspond à une décision en faveur de ω_k . Chaque colonne l correspond au cas où la vérité est ω_l . Le terme général $n_{k\ell}$ est égal au nombre d'objets de classe ω_l ayant été classés dans la classe ω_k par le classifieur.

La décision de « Rejet » correspond au cas où le classifieur ne discrimine aucun élément de l'univers : tous sont candidats pour être la véritable classe de l'objet en cours d'examen. Ainsi, un rejet complet est identifié à une décision en faveur de l'univers Ω . Le classifieur peut donc fournir $K + 1$ décisions possibles au regard d'un objet, et la vérité de chaque objet appartient à l'un des K éléments de Ω . Au total, la matrice de confusion d'un classifieur est de dimension $(K + 1, K)$ (Tableau 2.2).

Uniquement dans ce paragraphe, dans un but de simplicité d'écriture des expressions, Ω est aussi noté ω_{K+1} .

Notons $n_{k\cdot} = \sum_{\ell=1}^K n_{k\ell}$ le nombre d'objets classés dans ω_k , où $k \in \{1, \dots, K+1\}$. Le terme $n_{(K+1)\cdot}$ est le nombre de rejets effectués par le classifieur.

Notons $n = \sum_{k=1}^{K+1} \sum_{\ell=1}^K n_{k\ell}$ le nombre total d'objets classés.

À partir d'une telle matrice de confusion, deux méthodes de construction de fonction de croyance peuvent être employées :

- l'affectation Bayésienne allouant de la masse aux éléments singletons de Ω ;
- l'affectation de Xu et al. [120] allouant de la masse à la classe décidée par le classifieur, à son complémentaire et à l'ignorance.

Tableau 2.2 – Illustration d'une matrice de confusion.

Décision	Vérité	ω_1	\dots	ω_K
	ω_1		n_{11}	\dots
\vdots		\vdots	\ddots	\vdots
ω_K		n_{K1}	\dots	n_{KK}
$\Omega = \omega_{K+1}$		$n_{(K+1)1}$	\dots	$n_{(K+1)K}$

Ces deux méthodes sont présentées dans les paragraphes suivants. Remarquons qu'il peut en exister d'autres.

2.3.2 Affectation Bayésienne

Quand un classifieur fournit une décision $d = \{\omega_k\}$, l'affectation Bayésienne consiste à attribuer à chaque $\omega_\ell \in \Omega$, une masse égale au rapport du nombre d'objets de classe ω_ℓ ayant été classés dans ω_k , divisé par le nombre total d'objets classés dans ω_k :

$$\forall \omega_\ell \in \Omega, m(\{\omega_\ell\}) = \frac{n_{k\ell}}{\sum_{j=1}^K n_{kj}} = \frac{n_{k\ell}}{n_k}. \quad (2.6)$$

Tableau 2.3 – Exemple d'une matrice de confusion d'un classifieur en charge de la reconnaissance des caractères numériques de 1 à 9. Il est possible d'observer 60 décisions correctes, 16 décisions en erreur, et 20 rejets.

		Vérité								
		1	2	3	4	5	6	7	8	9
<i>D</i>	1	10	0	0	0	0	0	0	0	0
	2	0	8	0	0	0	0	0	0	0
<i>E</i>	3	0	0	3	0	0	0	0	0	0
<i>C</i>	4	0	0	0	4	0	0	0	0	0
<i>I</i>	5	0	0	0	0	5	0	0	0	0
<i>S</i>	6	0	0	0	0	0	6	0	2	0
<i>I</i>	7	0	4	0	0	0	0	7	0	0
<i>O</i>	8	0	0	0	0	0	0	0	8	4
<i>N</i>	9	0	0	2	4	0	0	0	0	9
	<i>Rejet</i>	1	2	3	4	5	2	3	0	0

Exemple 2.2. *Considérons un classifieur en charge de la reconnaissance des caractères numériques, dont la matrice de confusion est illustrée dans le tableau 2.3. En fonction de la décision d du classifieur, l'emploi de cette méthode conduit aux affectations suivantes.*

- $d = \{1\}$ implique $m(\{1\}) = 1$.
- $d = \{9\}$ implique $m(\{3\}) = 2/15$, $m(\{4\}) = 4/15$, $m(\{9\}) = 9/15$.
- $d = \text{« Rejet »}$ implique $m(\{1\}) = 1/20$, $m(\{2\}) = 2/20$, $m(\{3\}) = 3/20$, $m(\{4\}) = 4/20$, $m(\{5\}) = 5/20$, $m(\{6\}) = 2/20$, $m(\{7\}) = 3/20$.

Une telle affectation conduit à des fonctions de masse Bayésiennes. En conséquence, pour que cette précision soit justifiée et que le rapport $n_{k\ell}/n_k$ ait un sens, l'ensemble d'apprentissage doit être bien choisi et suffisamment grand. En particulier, quand le nombre de classes devient trop grand et qu'il n'est plus possible d'obtenir un ensemble d'apprentissage de cardinalité suffisante, il devient préférable d'exploiter des performances sur un groupe de classes, plutôt que sur des classes individuelles représentées par trop peu d'exemples.

2.3.3 Affectation de Xu et al. [120]

Les *taux de performance* de reconnaissance des classes de Ω peuvent être formellement définis de la manière suivante :

- le *taux de lecture* ou *de reconnaissance* R est égal à

$$R = \frac{\sum_{k=1}^K n_{kk}}{n}, \quad (2.7)$$

- le *taux d'erreur* ou *de substitution* S est égal à

$$S = \frac{\sum_{k=1}^K \sum_{\ell=1; \ell \neq k}^K n_{k\ell}}{n}, \quad (2.8)$$

- le *taux de rejet* T est égal à

$$T = \frac{n_{(K+1)}}{n}, \quad (2.9)$$

et, $R + S + T = 1$.

Quand un classifieur fournit une décision $d = \{\omega_k\}$ avec $k \in \{1, \dots, K\}$, il est proposé dans [120] d'utiliser ces trois taux pour définir m de la façon suivante :

$$\begin{array}{rcl} m : & 2^\Omega & \longrightarrow [0, 1] \\ & \{\omega_k\} & \longmapsto R \\ & \Omega \setminus \{\omega_k\} & \longmapsto S \\ & \Omega & \longmapsto T \end{array} \quad (2.10)$$

et, si $d = \{\omega_{K+1}\} = \text{« Rejet »}$, $m(\Omega) = 1$.

Exemple 2.3. *En reprenant le classifieur associé à la matrice de confusion illustrée sur le tableau 2.3. L'utilisation de la méthode Xu et al. conduit aux affectations suivantes.*

- $d = \{1\}$ implique $m(\{1\}) = 60/96$, $m(\Omega \setminus \{1\}) = 16/96$, $m(\Omega) = 20/96$.

- $d = \{9\}$ implique $m(\{9\}) = 60/96$, $m(\Omega \setminus \{9\}) = 16/96$, $m(\Omega) = 20/96$.
- $d = \text{« Rejet »}$ implique $m(\Omega) = 1$.

Dans ce cas, l'ajout par exemple d'une erreur sur la reconnaissance d'un 7 classé en un 1, ne change pas significativement le résultat de cette affectation, contrairement à l'affectation Bayésienne.

Cette méthode d'affectation est basée sur l'idée que plus le taux de reconnaissance est élevé, plus grande est la confiance dans la décision fournie par le classifieur. Cette affectation a été testée avec succès dans [120] sur un problème de reconnaissance de caractères numériques, où tous les classifieurs possédaient des taux de lecture comparables. Cependant, comme l'illustre l'exemple suivant, la part de croyance dans la décision du classifieur ne devrait pas dépendre seulement du taux de lecture.

Exemple 2.4. *Considérons deux classifieurs C_1 et C_2 avec les taux de performance suivants :*

	R	S	T
C_1	90%	1%	9%
C_2	20%	0.1%	79,9%

Supposons que C_1 décide $\{\omega_k\}$ et C_2 décide $\{\omega_{k'}\}$, avec $k \neq k'$. L'affectation de Xu et al. (2.10) conduit à $m_1(\{\omega_k\}) = 0.9$ et $m_2(\{\omega_{k'}\}) = 0.2$. Or, les décisions du classifieur C_2 différentes d'un rejet sont correctes la plupart du temps, ce qui n'est pas le cas des décisions du classifieur C_1 . Ainsi, cette affectation ne correspond pas à ce que nous pourrions attendre. Connaissant uniquement les taux R , S et T pour chaque classifieur, notre confiance dans les réponses de C_2 devrait être plus élevée que notre confiance en faveur des réponses de C_1 , la masse $m_2(\{\omega_{k'}\})$ devrait être plus élevée que $m_1(\{\omega_k\})$.

Aussi, nous proposons d'utiliser un autre taux de performance : le *taux de fiabilité*, afin de construire une méthode d'affectation voisine de celle de Xu et al.

2.3.4 Une affectation basée sur le taux de fiabilité

Le *taux de fiabilité* F d'un classifieur est défini par

$$F = \frac{\sum_{k=1}^K n_{kk}}{\sum_{k=1}^K \sum_{\ell=1}^K n_{k\ell}} = \frac{R}{1 - T}. \quad (2.11)$$

Ce taux mesure le pourcentage de bonnes réponses sans tenir compte des rejets. Il semble refléter, mieux que le taux de lecture, la fiabilité d'un classifieur lorsque celui-ci a pris une décision autre qu'un rejet.

Quand un classifieur décide $\{\omega_k\}$ avec $k \in \{1, \dots, K\}$, nous proposons alors l'affectation suivante :

$$\begin{aligned} m : \quad 2^\Omega &\longrightarrow [0, 1] \\ \{\omega_k\} &\longmapsto F \\ \Omega &\longmapsto 1 - F. \end{aligned} \quad (2.12)$$

De même que l'affectation de Xu et al. (2.10), quand un rejet est décidé, $m(\Omega) = 1$.

L'affectation basée sur le taux de fiabilité (2.12) prend en compte la fiabilité des classifieurs uniquement quand une classe est choisie.

Exemple 2.5. Appliquée à l'exemple 2.4, cette affectation conduit à $m_1(\{\omega_k\}) = \frac{0.9}{1-0.09} = 0.989$ et $m_2(\{\omega_{k'}\}) = \frac{0.2}{1-0.799} = 0.995$.

Cette affectation est un cas particulier des affectations relatives à des décisions exprimées sur une hiérarchie [64, 65], qui seront présentées dans le chapitre suivant.

2.4 Manipulation des connaissances

2.4.1 Révision versus mise à jour

Une révision des croyances se distingue d'une mise à jour des croyances [113]. Une *révision* consiste à élaborer une nouvelle croyance à partir des connaissances anciennes et nouvelles : les connaissances nouvelles viennent corriger, compléter, ou renforcer les anciennes. Au contraire, une *mise à jour* prend en compte l'évolution du monde dans lequel les croyances sont définies : toutes les croyances sont modifiées en accord avec les changements du monde [51] (voir aussi Levi [61], et Dubois et Prade [38]).

Dans le processus de fusion abordé dans ce mémoire, l'état du monde est fixé, et seules nos connaissances sur celui-ci évoluent. Ainsi, la révision de connaissances constitue le principal objet de ce paragraphe de manipulation des connaissances.

2.4.2 Le principe de minimum d'information

Il arrive fréquemment qu'une fonction de croyance, non entièrement définie, ne soit connue que par son appartenance à une famille de fonctions. Par exemple, cette famille peut décrire l'ensemble des fonctions de croyance compatibles avec une information. Le *principe de minimum d'information*³ consiste alors à choisir la fonction de croyance la moins *informative*, autrement dit, la moins *engagée*, présente dans cette famille.

Ce principe correspond à une forme de scepticisme, de conservatisme dans le transfert des masses, qui doivent être allouées aux sous-ensembles les plus larges possibles. Ce principe joue le même rôle que le principe de maximum d'entropie en probabilités [36, 50, 98].

Pour définir la notion d'informativité d'une fonction de croyance, deux approches sont possibles :

- une approche quantitative avec la définition de « mesures d'incertitude », par exemple la mesure de nonspécificité [32] définie par :

$$N(m) = \sum_{\emptyset \neq A \subseteq \Omega} m(A) \log_2(|A|) \quad (2.13)$$

- et, une approche ordinale permettant la construction d'un ordre partiel sur l'ensemble des fonctions de croyance [35, 122].

Dans le MCT, la définition d'informativité peut s'exprimer par une approche ordinale basée sur la fonction de plausibilité.

³En anglais, le terme *principle of minimal commitment* est employé.

Définition 2.8. Une fonction de croyance liée à une fonction de plausibilité pl_2 est dite moins informative qu'une fonction de croyance liée à une fonction de plausibilité pl_1 si et seulement si :

$$pl_1(A) \leq pl_2(A) \quad \forall A \subseteq \Omega. \quad (2.14)$$

Ainsi, ce principe pourrait aussi se nommer principe de maximum de plausibilité. Parmi toutes les fonctions de croyance définies sur Ω , la fonction la moins informative est la fonction de masse vide m_Ω .

2.4.3 L'idée de transfert des masses

Considérons une masse $m(A)$ strictement positive allouée à un sous-ensemble A de Ω . Si nous apprenons qu'avec certitude la vérité se situe dans un sous-ensemble B de Ω , alors la masse initialement allouée à A est transférée à $A \cap B$. Cette règle correspond à la règle de conditionnement non normalisée de Dempster. Le terme de « modèle des croyances transférables » est issu de cette idée de transfert des masses s'allouant à des sous-ensembles de l'univers en fonction des informations disponibles.

Outre cette règle de conditionnement, de nombreuses opérations sont possibles pour manipuler l'information représentée par les fonctions de croyance.

Toutes ces opérations peuvent être dérivées des notions de spécialisation et de généralisation formalisant ces transferts de masses vers des éléments plus restreints, ou au contraire plus grossiers.

2.4.4 Spécialisation et généralisation

Pour toutes fonctions de masses m_1 et m_2 appartenant à \mathcal{M}^Ω , il existe une matrice $\mathbf{M} = [M(A, B)]$, $A, B \subseteq \Omega$, transformant m_1 en m_2 , c'est-à-dire, telle que $\mathbf{m}_2 = \mathbf{M} \cdot \mathbf{m}_1$. En effet, en notant par $\mathbf{1}^t$, la transposée du vecteur de longueur $2^{|\Omega|}$ composé uniquement de 1, $\mathbf{1}^t \cdot \mathbf{m}_1$ est égale à la somme des masses de m_1 , c'est-à-dire 1, d'où une solution donnée par $\mathbf{M} = \mathbf{m}_2 \cdot \mathbf{1}^t$.

Plus précisément, Smets [107] montre le résultat suivant.

Théorème 2.1. L'ensemble des matrices associant à un élément de \mathcal{M}^Ω , un élément de \mathcal{M}^Ω est égal à l'ensemble des matrices $\mathbf{M} = [M(A, B)]$, $A, B \subseteq \Omega$, positives, carrées d'ordre $2^{|\Omega|}$, dont la somme sur chaque colonne est égale à 1 : $\sum_A M(A, B) = 1, \forall B$. Ces matrices sont appelées matrices stochastiques.

Preuve : Soit \mathbf{M} une matrice stochastique. Pour tout $\mathbf{m} \in \mathcal{M}^\Omega$, tous les éléments de $\mathbf{M} \cdot \mathbf{m}$ sont non négatifs et la somme de ces éléments est égale à $\mathbf{1}^t \cdot \mathbf{M} \cdot \mathbf{m} = \mathbf{1}^t \cdot \mathbf{m} = 1$. Donc $\mathbf{M} \cdot \mathbf{m} \in \mathcal{M}^\Omega, \forall \mathbf{m} \in \mathcal{M}^\Omega$.

Si \mathbf{M} n'est pas une matrice stochastique, alors il existe une colonne de \mathbf{M} , associée à un sous-ensemble A de Ω , qui contient une valeur négative ou dont la somme des éléments est différente de 1. Or, cette colonne est le résultat de l'application de \mathbf{M} à \mathbf{m}_A . Ainsi, $\mathbf{M} \cdot \mathbf{m}_A \notin \mathcal{M}^\Omega$. □

Lorsqu'une matrice stochastique $\mathbf{M} = [M(A, B)]$, $A, B \subseteq \Omega$, est appliquée à un vecteur de masses \mathbf{m} , le nombre $M(A, B)$ représente la fraction de la masse $m(B)$ qui est transférée à A .

La notion de spécialisation, déjà introduite par Kruse et Schwecke [57], Yager [121, 122]), et Dubois et Prade [35], ces derniers l'appelant inclusion forte, peut être définie, comme l'indiquent Klawonn et Smets [54, 107], de la manière suivante.

Définition 2.9. (*Spécialisation*) La fonction de masse m_1 est une spécialisation de m_2 si et seulement si, il existe $\mathbf{S} = [S(A, B)]$ une matrice stochastique vérifiant $S(A, B) = 0, \forall A \not\subseteq B$, telle que $\mathbf{m}_1 = \mathbf{S} \cdot \mathbf{m}_2$. La matrice \mathbf{S} est appelée matrice de spécialisation.

On vérifie que \mathbf{S} est une matrice triangulaire supérieure. Pour tout $B \subseteq \Omega$, la masse $m_2(B)$ est allouée à des sous-ensembles de B : la masse est transférée vers le bas. Rappelons que $S(A, B)$ représente la fraction de la masse $m_2(B)$ qui est transférée à A .

Exemple 2.6. Exemple d'une matrice de spécialisation \mathbf{S} avec $\Omega = \{a, b, c\}$. L'absence de valeur dans la matrice indique des valeurs devant être nulle pour que \mathbf{S} soit une matrice de spécialisation.

$$\mathbf{m}_1 = \mathbf{S} \cdot \mathbf{m}_2$$

	\mathbf{m}_1	=	\mathbf{S}	·	\mathbf{m}_2	
			$\begin{matrix} & \emptyset & a & b & a, b & c & a, c & b, c & a, b, c \end{matrix}$			
\emptyset	$\left(\begin{matrix} .13 \\ .04 \\ .13 \\ .00 \\ .50 \\ .00 \\ .12 \\ .08 \end{matrix} \right)$	=	$\left(\begin{matrix} 1 & .3 & .5 & .2 & .0 & .4 & .0 & .2 \\ & .7 & .5 & & .1 & & .1 & \\ & & .5 & .0 & & .4 & .0 & \\ & & & .3 & & & .0 & \\ & & & & 1 & .3 & .0 & .5 \\ & & & & & .2 & & .0 \\ & & & & & & .6 & .0 \\ & & & & & & & .2 \end{matrix} \right)$	$\left(\begin{matrix} .0 \\ .0 \\ .1 \\ .0 \\ .3 \\ .0 \\ .2 \\ .4 \end{matrix} \right)$		$\begin{matrix} \emptyset \\ a \\ b \\ a, b \\ c \\ a, c \\ b, c \\ a, b, c \end{matrix}$

D'autres exemples de matrices de spécialisation seront donnés par les matrices associées à la règle de conditionnement non normalisée de Dempster, et à la règle de combinaison conjonctive, encore appelée règle de combinaison non normalisée de Dempster.

Proposition 2.1. Si m_1 est une spécialisation de m_2 , alors m_1 est plus informative que m_2 .

Cette proposition a été prouvée par Dubois et Prade dans [35], par Yager dans [122]. Dans [107], Smets en fournit une démonstration plus synthétique en s'appuyant sur le calcul matriciel.

La notion de généralisation, notion duale de la spécialisation, peut être définie de la manière suivante.

Définition 2.10. (*Généralisation*) La fonction de masse m_2 est une généralisation de m_1 , si et seulement si, il existe $\mathbf{G} = [G(A, B)]$, $A, B \subseteq \Omega$, une matrice stochastique vérifiant $G(A, B) = 0$, pour tout A et B tels que $B \not\subseteq A$, telle que $\mathbf{m}_2 = \mathbf{G} \cdot \mathbf{m}_1$. La matrice \mathbf{G} est appelée matrice de généralisation.

Cette matrice est triangulaire inférieure : pour tout sous-ensemble B de Ω , la masse $m_1(B)$ sera transférée vers des éléments de Ω incluant B , autrement dit la masse remonte vers Ω .

Des exemples de matrice de généralisation seront donnés par les matrices associées à un déconditionnement, à un affaiblissement, ainsi qu'à nos travaux de généralisation de l'affaiblissement présentés au chapitre 4 [70, 71].

2.4.5 Connaissances distinctes

Définition 2.11 (Connaissances distinctes). *Considérons un agent Ag dont la connaissance est représentée par une fonction de masse m . Supposons qu'une nouvelle connaissance Ev soit connue par Ag . L'agent Ag doit alors réviser ses connaissances, c'est-à-dire, définir $m[Ev]$. Par le théorème 2.1, il existe une matrice stochastique $\mathcal{M}(Ev, \mathbf{m})$ transformant \mathbf{m} en $\mathbf{m}[Ev]$:*

$$\mathbf{m}[Ev] = \mathcal{M}(Ev, \mathbf{m}) \cdot \mathbf{m} . \quad (2.15)$$

Si $\mathcal{M}(Ev, \mathbf{m})$ ne dépend pas de \mathbf{m} , alors la connaissance ayant induit \mathbf{m} et la connaissance Ev seront dites distinctes, ou non interactives [105]. Dans ce cas,

$$\mathbf{m}[Ev] = \mathcal{M}(Ev) \cdot \mathbf{m}. \quad (2.16)$$

Une autre formulation de cette définition 2.11 est possible [96, 103]. Soient m_{Ag_1} et m_{Ag_2} deux fonctions de masse induites respectivement par Ev_1 et Ev_2 . Si le choix de m_{Ag_1} par Ag_1 pour représenter sa connaissance implique une restriction dans le choix de Ag_2 pour représenter sa connaissance, c'est-à-dire si Ag_2 ne peut pas choisir sa fonction de masse dans \mathcal{M}^Ω tout entier, alors Ev_1 et Ev_2 ne sont pas distinctes.

En pratique, deux connaissances sont considérées comme distinctes si les sources associées à ces connaissances ne sont aucunement « liées ».

2.4.6 Conditionnement et déconditionnement

Le conditionnement d'une fonction de masse m^Ω (paragraphe 2.4.3) sur un sous-ensemble B de Ω peut être défini comme la spécialisation de m^Ω la moins informative telle que $pl[B](\overline{B}) = 0$ (Klawonn et Smets [54]). Le conditionnement de m^Ω par B est noté $m^\Omega[B]$ et vérifie :

$$m^\Omega[B](A) = \begin{cases} \sum_{X: X \subseteq \overline{B}} m(A \cup X) & \text{si } A \subseteq B \\ 0 & \text{sinon.} \end{cases} \quad (2.17)$$

La matrice de spécialisation \mathbf{C}_B associée à un conditionnement sur B est définie par :

$$\mathbf{C}_B(A, C) = \begin{cases} 1 & \text{si } A = B \cap C \\ 0 & \text{sinon.} \end{cases} \quad (2.18)$$

On vérifie que $\mathbf{m}[B] = \mathbf{C}_B \cdot \mathbf{m}$.

Exemple 2.7. Exemple d'un conditionnement sur $B = \{a, c\}$ d'une fonction de masse m avec $\Omega = \{a, b, c\}$.

$$\begin{array}{l}
\mathbf{m}[B] \\
\emptyset \\
a \\
b \\
a, b \\
c \\
a, c \\
b, c \\
a, b, c
\end{array}
=
\begin{array}{c}
\mathbf{C}_B \\
\emptyset \quad a \quad b \quad a, b \quad c \quad a, c \quad b, c \quad a, b, c \\
\left(\begin{array}{cccccccc}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
& 1 & & 1 & & 0 & & \\
& & 0 & 0 & & & 0 & 0 \\
& & & 0 & & & & 0 \\
& & & & 1 & 0 & 1 & 0 \\
& & & & & 1 & & 1 \\
& & & & & & 0 & 0 \\
& & & & & & & 0
\end{array} \right)
\end{array}
=
\begin{array}{c}
\mathbf{m} \\
\left(\begin{array}{c}
.0 \\
.0 \\
.1 \\
.0 \\
.3 \\
.0 \\
.2 \\
.4
\end{array} \right)
\end{array}
\begin{array}{l}
\emptyset \\
a \\
b \\
a, b \\
c \\
a, c \\
b, c \\
a, b, c
\end{array}$$

Le nombre $\mathbf{C}_B(A, C)$ représente la fraction de la masse $m(C)$ qui est transférée à A lors d'un conditionnement sur B . Pour tout $C \subseteq \Omega$, la masse $m(C)$ est transférée sur $A = B \cap C$.

Le déconditionnement est l'opération duale du conditionnement. Étant donnée une fonction de masse conditionnelle $m[B]$, il n'est généralement pas possible de retrouver la fonction de masse originelle m . Cependant, la fonction de masse la moins informative qui, conditionnée sur B , donnera $m[B]$, peut être calculée. Celle-ci est notée $m[B]^{\uparrow\Omega}$, et est définie par $m[B]^{\uparrow\Omega}(A \cup \bar{B}) = m[B](A)$, $\forall A \subseteq B$. La matrice de généralisation \mathbf{D}_B associée à un déconditionnement par rapport à B est définie par :

$$\mathbf{D}_B(A, C) = \begin{cases} 1 & \text{si } A = \bar{B} \cup C \\ 0 & \text{sinon.} \end{cases} \quad (2.19)$$

La matrice de généralisation \mathbf{D}_B associée à un déconditionnement par rapport à $B = \{a, c\}$ avec $\Omega = \{a, b, c\}$, est illustrée dans l'exemple suivant.

Exemple 2.8. Exemple de déconditionnement par rapport à $B = \{a, c\}$ d'une fonction de masse $m[B]$ avec $\Omega = \{a, b, c\}$.

$$\begin{array}{l}
\mathbf{m}[B]^{\uparrow} \\
\emptyset \\
a \\
b \\
a, b \\
c \\
a, c \\
b, c \\
a, b, c
\end{array}
=
\begin{array}{c}
\mathbf{D}_B \\
\emptyset \quad a \quad b \quad a, b \quad c \quad a, c \quad b, c \quad a, b, c \\
\left(\begin{array}{cccccccc}
0 & & & & & & & \\
0 & 0 & & & & & & \\
1 & & 1 & & & & & \\
0 & 1 & 0 & 1 & & & & \\
0 & & & & 0 & & & \\
0 & 0 & & & 0 & 0 & & \\
0 & & 0 & & 1 & & 1 & \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array} \right)
\end{array}
=
\begin{array}{c}
\mathbf{m}[B] \\
\left(\begin{array}{c}
.1 \\
.0 \\
.0 \\
.0 \\
.5 \\
.4 \\
.0 \\
.0
\end{array} \right)
\end{array}
\begin{array}{l}
\emptyset \\
a \\
b \\
a, b \\
c \\
a, c \\
b, c \\
a, b, c
\end{array}$$

Le nombre $\mathbf{D}_B(A, C)$ représente la fraction de la masse $m[B](C)$ qui est transférée à A lors d'un déconditionnement par rapport à B . Pour tout $C \subseteq \Omega$, la masse $m[B](C)$ est transférée sur $A = \bar{B} \cup C = \{b\} \cup C$.

2.4.7 Combinaisons conjonctive et disjonctive

Définition 2.12 (Matrices de spécialisation Dempstériennes). Soit $m \in \mathcal{M}^\Omega$. La matrice de spécialisation \mathbf{S}_m dont les coefficients vérifient :

$$S_m(A, B) = m[B](A), \quad \forall A, B \subseteq \Omega, \quad (2.20)$$

est appelée matrice de spécialisation Dempstérienne associée à m .

Exemple 2.9. Exemple d'une matrice de spécialisation Dempstérienne \mathbf{S}_m associée à une fonction de masse m , avec $\Omega = \{a, b, c\}$.

$$\begin{array}{c} \mathbf{m} \\ \left(\begin{array}{c} .0 \\ .0 \\ .1 \\ .0 \\ .3 \\ .0 \\ .2 \\ .4 \end{array} \right) \end{array} \quad \begin{array}{c} \mathbf{S}_m \\ \left(\begin{array}{cccccccc} \emptyset & a & b & a, b & c & a, c & b, c & a, b, c \\ 1 & .6 & .3 & .3 & .1 & .1 & .0 & .0 \\ & .4 & & .0 & & .0 & & .0 \\ & & .7 & .3 & & & .1 & .1 \\ & & & .4 & & & & .0 \\ & & & & .9 & .5 & .3 & .3 \\ & & & & & .4 & & .0 \\ & & & & & & .6 & .2 \\ & & & & & & & .4 \end{array} \right) \end{array}$$

Proposition 2.2. Pour tout $m_1, m_2 \in \mathcal{M}^\Omega$, $\mathbf{S}_{m_1} \cdot \mathbf{m}_2 = \mathbf{m}_1 \circledast_2$, où $\mathbf{m}_1 \circledast_2$ désigne la combinaison de Dempster non normalisée, ou combinaison conjonctive de m_1 et m_2 , définie par :

$$m_1 \circledast_2(A) = m_1 \circledast m_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega. \quad (2.21)$$

Preuve : Pour tout $A \subseteq \Omega$:

$$\begin{aligned} \mathbf{S}_{m_1} \cdot \mathbf{m}_2(A) &= \sum_{B \subseteq \Omega} S_{m_1}(A, B) m_2(B) \\ &= \sum_{B \subseteq \Omega} m_1[B](A) m_2(B) \\ &= \sum_{B \subseteq \Omega} \left[\sum_{C \subseteq \Omega} C_B(A, C) m_1(C) \right] m_2(B) \\ &= \sum_{B \subseteq \Omega} \left[\sum_{C \subseteq \Omega: A=B \cap C} m_1(C) \right] m_2(B) \\ &= \sum_{B, C \subseteq \Omega: A=B \cap C} m_1(C) m_2(B) \\ &= \mathbf{m}_1 \circledast_2(A). \end{aligned}$$

□

En outre, Klawonn et Smets [54] ont montré le théorème suivant.

Théorème 2.2 (Klawonn et Smets, 1992, [54]). *L'ensemble le plus large des matrices de spécialisation qui commutent et contenant les matrices associées à un conditionnement, est égal à l'ensemble des matrices de spécialisation Dempstériennes.*

Ainsi, la combinaison conjonctive dérive uniquement des notions de spécialisation, associativité, commutativité, et du principe selon lequel le conditionnement est une forme particulière de combinaison conjonctive. Remarquons enfin que ce conditionnement peut lui-même, comme mentionné en début de paragraphe 2.4.6, se définir à partir de la notion de spécialisation la moins informative telle que $pl[A](\overline{A}) = 0$, avec A un sous-ensemble de Ω .

D'autres justifications axiomatiques de la règle de combinaison de Dempster ont été proposées [34, 48, 53, 95]. Cependant, la justification présentée dans ce paragraphe nous semble la plus simple et la plus convaincante.

L'associativité et la commutativité assurent que l'ordre dans lequel les sources sont combinées n'a pas d'importance, ce qui est naturellement requis : lorsque plusieurs informations sont à fusionner, il serait critiquable de devoir définir un ordre artificiel pour s'assurer de l'unicité du résultat.

Remarque 2.3. *L'application de \mathbf{S}_m à la fonction de masse vide, induite par une ignorance totale, conduit à la fonction de masse $\mathbf{m} : \mathbf{S}_m \cdot \mathbf{m}_\Omega = \mathbf{m}$, qui est induite par Ev . Ainsi, la matrice Dempstérienne \mathbf{S}_m caractérise l'impact de l'information Ev ayant induit m .*

Remarque 2.4. *Les éléments présents sur la diagonale de \mathbf{S}_m sont égaux aux communalités de m . En effet, pour tout $A \subseteq \Omega$, $q(A) = mA$.*

Enfin, le conditionnement d'une fonction de masse sur B peut être donné par l'expression suivante :

$$m[B] = m \odot m_B. \quad (2.22)$$

En effet, $\mathbf{m}[B] = \mathbf{C}_B \cdot \mathbf{m} = \mathbf{S}_{m_B} \cdot \mathbf{m} = \mathbf{m} \odot \mathbf{m}_B$.

De manière similaire, la combinaison disjonctive peut être obtenue à partir des notions de généralisation et de déconditionnement. Pour cela, il suffit de définir une matrice de généralisation \mathbf{G}_m , $m \in \mathcal{M}^\Omega$, dont les coefficients vérifient $G_m(A, B) = m[\overline{B}]^\dagger(A)$, pour tout $A, B \subseteq \Omega$. Pour tout $m_1, m_2 \in \mathcal{M}^\Omega$, on vérifie que $\mathbf{G}_{m_1} \cdot \mathbf{m}_2 = \mathbf{m}_1 \odot m_2$, où $m_1 \odot m_2$ désigne la combinaison disjonctive de m_1 et m_2 , définie par :

$$m_1 \odot m_2(A) = \sum_{B \cup C = A} m_1(B)m_2(C), \quad \forall A \subseteq \Omega, \quad m_1, m_2 \in \mathcal{M}^\Omega. \quad (2.23)$$

D'un point de vue pratique, si $m_1 = m[Ev_1]$, et $m_2 = m[Ev_2]$ sont deux fonctions de masse induites par les deux connaissances distinctes Ev_1 et Ev_2 , alors la combinaison conjonctive consiste à accepter ces deux informations :

$$m[Ev_1] \odot m[Ev_2] = m[Ev_1 \wedge Ev_2] = \sum_{B \cap C = A} m_1(B)m_2(C), \quad \forall A \subseteq \Omega. \quad (2.24)$$

et, la combinaison disjonctive consiste à ne prendre en compte que l'une ou l'autre information :

$$m[Ev_1] \odot m[Ev_2] = m[Ev_1 \vee Ev_2] = \sum_{B \cup C = A} m_1(B)m_2(C), \quad \forall A \subseteq \Omega, \quad (2.25)$$

Remarque 2.5. Outre cette conjonction et disjonction des connaissances, d'autres opérateurs logiques peuvent être employés, comme par exemple le ou-exclusif et la négation (Dubois et Prade [35]). Une généralisation, où ces opérations peuvent être pondérées, a été proposée par Smets [102].

Les combinaisons conjonctive et disjonctive peuvent être simplement exprimées par les fonctions de communalité et d'implicabilité :

- la fonction de communalité $q_1 \odot_2$ associée à $m_1 \odot m_2$ peut être obtenue par le produit de q_1 et q_2 :

$$q_1 \odot_2(A) = q_1(A)q_2(A), \quad \forall A \subseteq \Omega, \text{ et,} \quad (2.26)$$

- la fonction d'implicabilité $b_1 \oplus_2$ associée à $m_1 \oplus m_2$ peut être obtenue par le produit de b_1 et b_2 :

$$b_1 \oplus_2(A) = b_1(A)b_2(A), \quad \forall A \subseteq \Omega. \quad (2.27)$$

Dans [107], en appliquant le calcul matriciel aux fonctions de croyance, Smets fournit une démonstration élégante de ces expressions. Afin de l'illustrer, la preuve de la première expression (2.26) est reprise dans ce mémoire.

En sachant que les valeurs propres d'une matrice de spécialisation Dempstérienne S_m sont les communalités associées à m , et que la matrice contenant les vecteurs propres de S_m est la matrice de transformation d'une fonction de masse en fonction de communalité $QfrM$ [54], on a :

$$\begin{aligned} & q_1 \odot_2 \\ = & QfrM \cdot m_1 \odot_2 && \text{(définition de } QfrM) \\ = & QfrM \cdot S_{m_1} \cdot m_2 && \text{(proposition 2.2)} \\ = & QfrM \cdot MfrQ \cdot Diag(q_1) \cdot QfrM \cdot m_2 && \text{(remarque ci-avant)} \\ = & Diag(q_1) \cdot QfrM \cdot m_2 \\ = & Diag(q_1) \cdot q_2, \end{aligned}$$

où $Diag(q_1)$ est la matrice diagonale composée des communalités de q_1 .

Ces expressions (2.26) et (2.27) illustrent l'intérêt des fonctions de communalité et d'implicabilité.

2.4.8 Affaiblissement

Il est parfois possible d'avoir un doute sur la fiabilité d'une information m fournie par une source S . À partir d'une constante $\alpha \in [0, 1]$, appelé *taux d'affaiblissement*, l'opération d'*affaiblissement* de m , définie par :

$$\begin{cases} \alpha m(A) &= (1 - \alpha)m(A), \quad \forall A \subset \Omega, \\ \alpha m(\Omega) &= (1 - \alpha)m(\Omega) + \alpha, \end{cases} \quad (2.28)$$

ou, plus simplement :

$$\alpha m = (1 - \alpha)m + \alpha m_\Omega, \quad (2.29)$$

permet de prendre en compte cette *métaconnaissance* relative à la fiabilité de l'information fournie par la source.

Le coefficient $(1 - \alpha)$ représente le degré de fiabilité de la source. Si la source n'est pas fiable, le degré de fiabilité de la source $(1 - \alpha)$ est nul, et ${}^\alpha m$ est égale à la fonction de masse vide. Au contraire, si la source est fiable, le degré de fiabilité $(1 - \alpha)$ est égal à 1, et m n'est aucunement affaiblie, le taux d'affaiblissement α est nul. En pratique, la fiabilité d'une source n'est souvent pas connue avec certitude, d'où un degré de fiabilité entre 0 et 1.

Comme Smets l'a montré [98], cette opération d'affaiblissement n'est pas *ad hoc*. Ce *mécanisme de correction de fonction de croyance* peut être obtenu à partir d'un simple modèle de fiabilité, qui sera exposé en détail dans le paragraphe 4.3 du chapitre 4.

2.4.9 Le degré de conflit $m(\emptyset)$

Lorsque deux sources sont combinées conjonctivement, la masse portant sur l'ensemble vide peut devenir strictement positive. Cette masse mesure alors le conflit entre les sources.

De même que l'ignorance [39, 46], cette mesure de conflit est très utile, et il convient de ne pas la masquer par une normalisation aveugle, comme l'illustre le célèbre exemple de Zadeh [125]. Ainsi, dans le MCT, cette masse est conservée et joue un rôle d'alarme [110] : plus cette masse est importante, plus l'adéquation entre le modèle et la réalité peut être mise en doute.

Comme le souligne Smets dans un article de synthèse sur la combinaison de fonctions de croyance conflictuelles [110], les origines possibles à la présence du conflit sont multiples et non exclusives :

1. Le cadre de discernement peut être non exhaustif : c'est l'hypothèse du monde ouvert de Smets [97]. L'objet à identifier n'est pas connu des sources combinées.
2. Une des hypothèses (2.24) de l'application de la combinaison conjonctive peut ne pas être satisfaite :
 - les sources ne se sont pas basées sur des connaissances distinctes, ou
 - au moins l'une des sources n'est pas fiable.
3. Les sources peuvent ne pas faire référence au même objet.

Dans le cas où la masse sur le conflit n'est pas acceptable, les différents traitements suivants, dépendant de l'application visée, peuvent être employés :

- l'exhaustivité du cadre de discernement peut être contrôlée par une fermeture artificielle : un élément associé à l'ensemble des objets inconnus peut être ajouté [78, 82] ;
- la non distinctivité des sources peut être prise en compte par l'emploi d'une combinaison plus prudente. Outre les combinaisons disjonctive, moyenne, de Dubois et Prade [35, 37], ou de Yager [123], la combinaison prudente introduite récemment par Dencœux dans [28] constitue un très bon candidat. Celle-ci est associative, commutative et idempotente, et constitue un cas particulier d'une famille de combinaison exploitant la décomposition canonique d'une fonction de masse introduite par Smets [101] ;
- la non fiabilité d'une source d'information peut être traitée par un affaiblissement ou une élimination de cette source. Si aucune information sur la fiabilité

des sources n'est disponible, un apprentissage automatique de ces coefficients peut être réalisé, à partir de données étiquetées, avec par exemple la méthode de Elouedi et al. [40], sur laquelle nous reviendrons un peu plus loin dans ce mémoire au paragraphe 4.6 du chapitre 4 ;

- enfin, le conflit peut être véritablement exploité comme une mesure de cohérence d'hypothèses réalisées sur le nombre d'objets et l'association entre les sources disponibles et ces objets. Par exemple, la minimisation du conflit entre les fonctions de croyance issues des sources en fonction de ces hypothèses, est à l'origine de méthodes de clustering (Schubert [84–86]), de détection d'objets (Ayoun et Smets [5]), ou de suivi de cibles (Ristic et Smets [75]).

À chaque itération du problème d'une application, le conflit peut avoir une origine différente et ainsi demander un traitement particulier. Aussi, il est proposé par Smets dans [110], de généraliser l'emploi d'un système expert, placé en fin de combinaison, et qui serait en charge d'assurer la gestion du conflit, comme par exemple cela est réalisé par Milisavljević et Bloch dans [72].

Remarque 2.6. *Outre ces raisons, d'une manière très pratique, l'origine de la présence d'un conflit peut se révéler plus profonde : le conflit peut être automatiquement créé par une mauvaise modélisation du problème traité. En effet, le choix des cadres de discernement sur lesquels sont modélisées les informations disponibles, et la méthode de construction de ces fonctions de croyance peuvent être difficiles à appréhender. Ainsi, lorsque le conflit s'avère omniprésent, la modélisation elle-même du problème devrait être « repensée ».*

2.4.10 Grossissements et raffinements

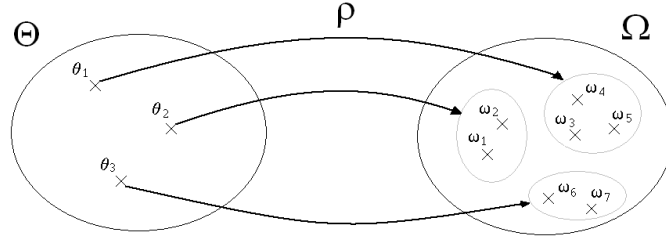
Dans le cadre du MCT, la détermination du cadre de discernement est une étape cruciale dans la modélisation de toute application réelle. Comme le souligne Shafer [88, chapitre 6], la détermination de ce cadre est toujours une affaire de « convention », étant donné que chaque élément du cadre de discernement représente un état du monde, qui peut toujours être divisé en différentes alternatives. Ainsi, les concepts de *grossissement* (« coarsening ») et *raffinement* (« refinement ») ont été définis de la manière suivante.

Étant donnés Θ et Ω deux cadres de discernement. Une fonction $\rho : 2^\Theta \rightarrow 2^\Omega$ est appelée *raffinement* de Θ si et seulement si :

1. l'ensemble $\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega$ est une partition de Ω , et
2. pour tout $A \subseteq \Theta$:

$$\rho(A) = \bigcup_{\theta \in A} \rho(\{\theta\}). \quad (2.30)$$

Dans ce cas, illustré sur la figure 2.2, Θ est appelé grossissement de Ω , et Ω est appelé raffinement de Θ .

Figure 2.2 – Illustration [27] d'un grossissement Θ de Ω , avec raffinement ρ de Θ .

2.5 Niveau pignistique

2.5.1 Transformation pignistique

Lorsqu'une décision doit être prise, des *principes de rationalité* [19, 74, 83] justifient la stratégie consistant à choisir la décision d parmi un ensemble de décisions possibles \mathcal{D} minimisant le *risque espéré* défini par :

$$\rho(d) = \sum_{\omega \in \Omega} c(d, \omega) P^\Omega(\{\omega\}), \quad (2.31)$$

où $P^\Omega : 2^\Omega \rightarrow [0, 1]$ est une mesure de probabilité et $c : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$ une fonction de coût, $c(d, \omega)$ représentant le coût de décider d sachant que la vérité est ω .

À ce niveau, la fonction de croyance obtenue après combinaison doit donc être *transformée* en une mesure de probabilité. Une solution consiste à utiliser la *transformation pignistique* en calculant la *probabilité pignistique* [24, 31, 108, 111] définie par :

$$\text{Bet}P^\Omega(\{\omega\}) = \sum_{\{A \subseteq \Omega, \omega \in A\}} \frac{m(A)}{|A| (1 - m(\emptyset))}. \quad (2.32)$$

La probabilité pignistique se justifie principalement à partir de l'axiome de linéarité de la transformation pignistique par rapport à la fonction de masse, qui lui-même peut se justifier à partir du principe de maximisation de l'utilité espérée [108]. La solution obtenue consiste à répartir uniformément la masse de chaque élément focal entre les singletons qui le composent.

2.5.2 Cadre de pari

Le cadre de pari Γ est le cadre sur lequel les décisions doivent être prises. Ce cadre peut être égal à Ω , comme dans l'équation (2.32) ou être différent. Il est construit à partir d'un grossissement ou raffinement de Ω .

La transformation pignistique d'une fonction de croyance bel^Ω dépend ainsi de bel^Ω et du cadre de pari Γ . L'exemple suivant, résumé de l'exemple du « gardien dans la centrale électrique » souvent repris par Smets, illustre cette dépendance.

Exemple 2.10. *Considérons un cadre de discernement $\Omega = \{\omega_1, \omega_2\}$ composé de deux éléments, et un second cadre $\Theta = \{\theta_1, \theta_2, \theta_3\}$ composé de trois éléments.*

Dans un premier temps, l'agent Ag n'a aucune information sur la valeur de la vérité dans Ω ou Θ . Ainsi $bel^\Omega(\Omega) = bel^\Theta(\Theta) = 1$, et $BetP^\Omega(\omega_k) = \frac{1}{2}, \forall k$, et $BetP^\Theta(\theta_\ell) = \frac{1}{3}, \forall \ell$.

Dans un second temps, l'agent apprend l'information suivante : la vérité dans Ω est ω_1 , si et seulement si, la vérité dans Θ est θ_1 . Dans le MCT, cette information ne modifie ni les croyances bel^Ω et bel^Θ , on a toujours $bel^\Omega(\omega_1) = bel^\Theta(\theta_1) = 0$, ni les probabilités pignistiques $betP^\Omega$ et $betP^\Theta$. La différence entre $BetP^\Omega$ et $BetP^\Theta$ reflète la différence de cadre de pari.

Remarquons que dans un cadre Bayésien, l'information obtenue dans le second temps entraîne que la probabilité sur ω_1 et la probabilité sur θ_1 doivent être égale, et ainsi des contradictions dans la détermination de cette valeur.

Ainsi, la probabilité pignistique est le résultat de la transformation d'une fonction de croyance en une mesure de probabilité, mais aussi de la transformation d'une fonction de croyance bel^Ω définie à partir du cadre de discernement Ω , en une fonction de croyance bel^Γ définie à partir du cadre de pari Γ .

2.5.3 Risques pignistique, inférieur et supérieur

L'utilisation de la probabilité pignistique aux principes de rationalité (2.31), conduit à une prise de décision basée sur la minimisation du risque pignistique définie par :

$$\rho_{bet}(d) = \sum_{\gamma \in \Gamma} c(d, \gamma) BetP^\Gamma(\{\gamma\}). \quad (2.33)$$

Outre ce risque, d'autres risques peuvent être employés [24], comme les risques inférieur et supérieur définis respectivement par :

$$\rho_*(d) = \sum_{A \subseteq \Gamma} m^\Gamma(A) \min_{\gamma_k \in A} c(d, \gamma_k), \quad (2.34)$$

et,

$$\rho^*(d) = \sum_{A \subseteq \Gamma} m^\Gamma(A) \max_{\gamma_k \in A} c(d, \gamma_k). \quad (2.35)$$

Le choix d'une décision de risque inférieur (resp. supérieur) minimal correspond à une stratégie optimiste (resp. pessimiste).

Exemple 2.11. *Dans le cas de coûts 0-1, c'est-à-dire un coût de 1 en cas d'erreur et un coût de 0 en cas de réponse correcte, la minimisation du risque pignistique (resp. inférieur, supérieur) est équivalente au choix de la décision de probabilité pignistique (resp. plausibilité, croyance) maximale. L'utilisation du maximum de plausibilité est par exemple utilisé par Appriou [2].*

Remarque 2.7. *Une prise de décision à partir d'une transformation pignistique basée sur la plausibilité (Cobb et Shenoy [15]), peut aussi se justifier à partir d'autres propriétés escomptées [15, 16, 109].*

2.6 Conclusion

Dans ce chapitre, les principaux concepts du MCT ont été exposés :

- une fonction de croyance est une opinion pondérée : à chaque alternative du monde est associé un nombre entre 0 et 1 ;
- deux niveaux sont distingués : un niveau crédal et un niveau pignistique ;
- au niveau crédal, seules des fonctions de croyance sont manipulées ;
- à tout moment, toutes les informations disponibles peuvent être prises en compte de par la souplesse de représentation du modèle, et de par l'existence de nombreuses opérations de transfert de croyances ;
- au niveau pignistique, lorsqu'une décision doit être prise, une fonction de probabilité est construite à partir de la fonction de croyance résultante de toutes les informations disponibles au niveau crédal.

Ce tour d'horizon a montré tout l'intérêt de ce modèle mathématique pour la fusion d'informations incertaines et imprécises.

Dans le prochain chapitre, un modèle de fusion de décisions postales basé sur le MCT est exposé.

Fusion de décisions postales dans le cadre du MCT

Dans ce chapitre, un modèle de fusion de décisions postales est présenté [64, 66, 67]. Le choix du cadre de discernement, une étape importante dans la construction de tout modèle utilisant des fonctions de croyance, est décrit dans le paragraphe suivant.

3.1 Choix du cadre de discernement

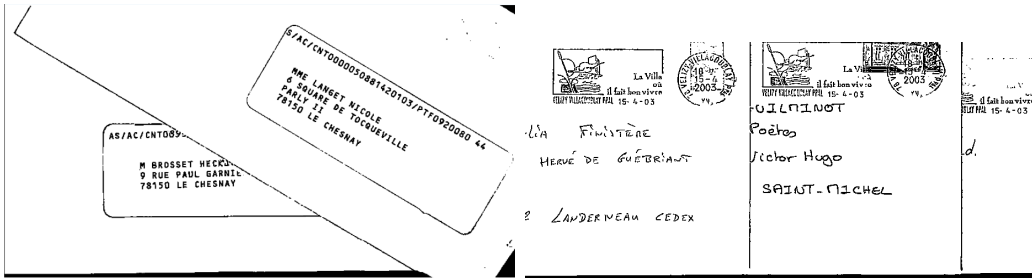
Notre application a pour but d'identifier l'adresse du destinataire d'un courrier postal. Cette adresse doit se trouver inscrite sur le courrier et son code doit se trouver dans la base de données.

Dans la réalité, il est possible que l'adresse inscrite sur ce courrier comporte des erreurs plus ou moins importantes. Ces erreurs peuvent être de natures très diverses (figure 3.1) :

- une contradiction peut être présente ; par exemple : le code postal inscrit ne correspond pas à l'intitulé de la ville, ou la rue n'est pas dans la ville ;
- une information peut être absente ; par exemple : le numéro de pas de porte ou le code postal est oublié, voire même, il n'y a pas d'adresse inscrite sur l'enveloppe ;
- des informations peuvent être indéchiffrables, raturées, illisibles même pour un humain ;
- l'adresse inscrite peut ne pas se trouver dans la base de données ou ne pas exister ;
- des cas de superposition ou de retournement de lettres peuvent aussi se produire.

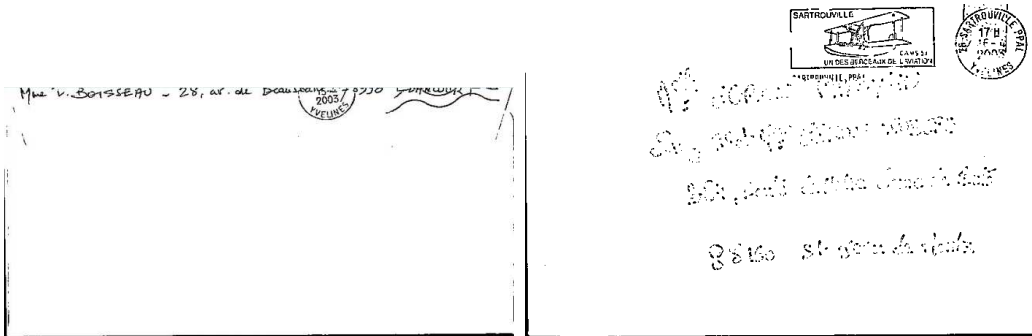
Chaque type d'erreur doit ou non être corrigé en fonction de règles spécifiques au pays traité. Les adresses ne contenant pas d'erreurs ou contenant une erreur devant être corrigée, seront dites *valides*. Les adresses comportant des erreurs non rattrapables seront dites :

- *invalides ou totalement invalides* dans le cas où la ville, qui constitue le minimum à reconnaître, ne peut pas être reconnue ;
- *partiellement invalides* lorsque l'intitulé de l'adresse complète ne peut pas être reconnu, mais il est possible de reconnaître des éléments de l'adresse situés au moins au niveau acheminement.



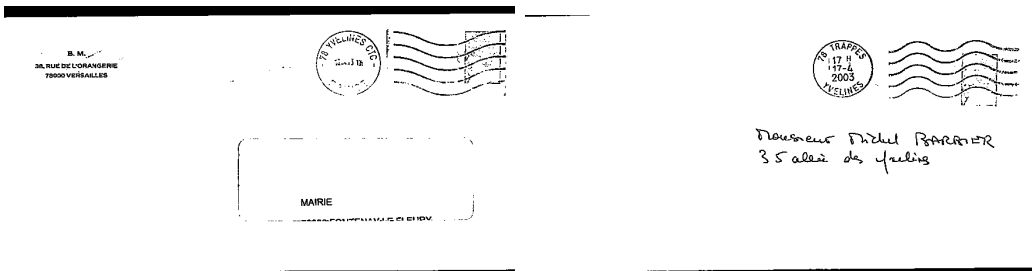
(a) Deux lettres superposées

(b) Autre superposition



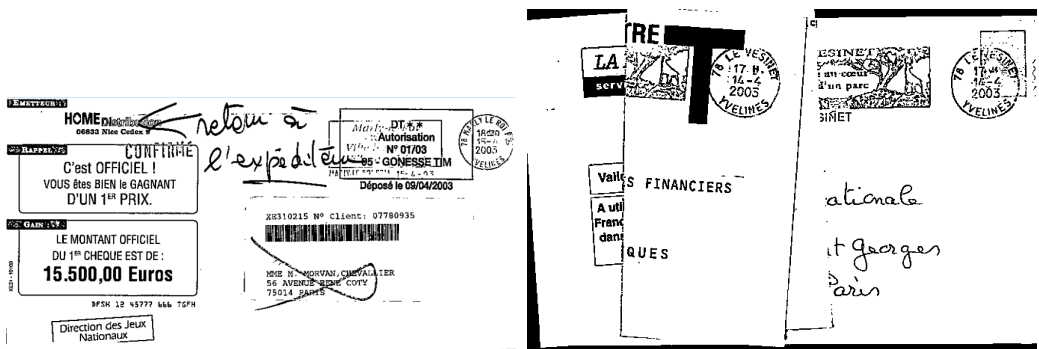
(c) Une lettre retournée

(d) Encre effacée



(e) Problème lié à une enveloppe à fenêtre. Un danger consiste ici à décider l'adresse de l'expéditeur qui est la seule adresse lisible sur cet envoi.

(f) Absence d'acheminement



(g) Une lettre dont le destinataire est considéré comme non reconnaissable

(h) Lettre pliée

Figure 3.1 – Des exemples de lettres dont l'adresse est invalide.

Avec les niveaux considérés sur la figure 3.2, les adresses partiellement invalides correspondent soit au cas où la ville est reconnaissable mais pas la rue, soit au cas où la ville et la rue sont reconnaissables mais pas le numéro de pas de porte.

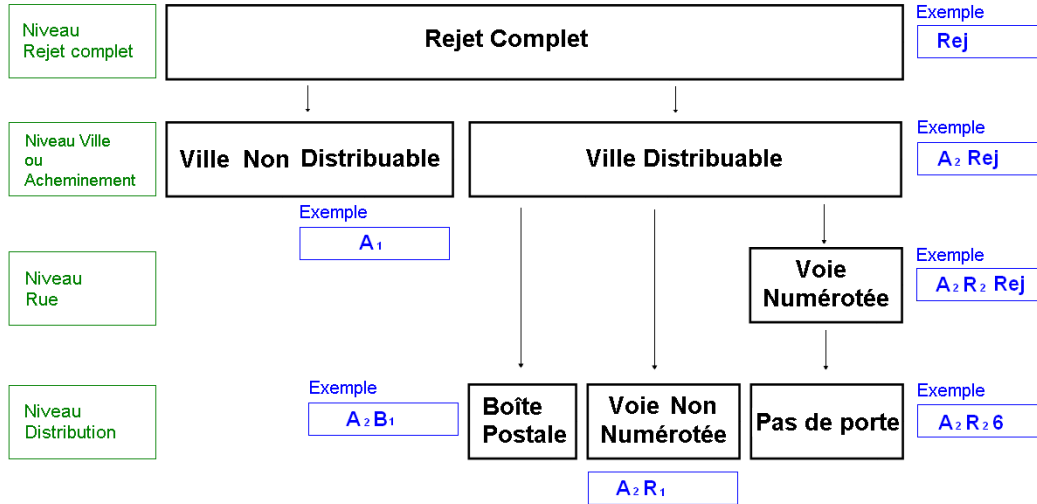


Figure 3.2 – Rappel de la hiérarchie de référence.

Ainsi, le cadre de discernement Ω , commun à toutes les fonctions de croyance de ce modèle, sera constitué de l'ensemble des adresses valides, de l'ensemble des adresses partiellement invalides et d'un élément représentant les adresses totalement invalides. Explicitement, les éléments singletons de Ω sont :

1. les adresses associées à une distribution géographique voie numérotée ;
2. les adresses associées à une distribution géographique voie non numérotée ;
3. les adresses associées à une distribution spéciale ;
4. les adresses associées à un acheminement non distribuable ;
5. une adresse totalement invalide, dont le code sera « inv » ;
6. les adresses partiellement invalides, dont la partie invalide sera noté par « inv ». Par exemple, si l'acheminement d'un envoi est « A_2 », mais la partie distribution est illisible ou absente, la vérité de cet envoi sera notée par le code « $A_2 inv$ ».

Une autre particularité de cette application réside dans le fait que les LAP ne formulent pas de décision en faveur des adresses partiellement ou totalement invalides. En présence d'incohérences sur un courrier invalide, ou de difficultés de reconnaissance sur un courrier valide, ils effectuent un rejet au niveau de la partie posant problème. Par exemple, si un LAP estime fortement que la vérité est « $A_2 R_2 inv$ », le numéro de pas de porte étant invalide, il va décider « $A_2 R_2 Rej$ ». Les rejets ont ainsi une double signification. Par le cadre de discernement proposé, cette ambiguïté est levée : soit l'adresse existe, même si elle est très difficile à reconnaître, soit l'adresse comporte une part d'invalidité.

Exemple 3.1 (Sur une base de données réduite). *Considérons une base de données ne contenant que deux acheminements, A_1 et A_2 , où A_1 est non distribuable et A_2 est distribuable. Le point d'acheminement A_2 est composé de deux boîtes postales, B_1 et B_2 , d'une rue non numérotée R_1 , et d'une rue numérotée R_2 comportant seulement trois numéros 2, 4 et 6. Dans ce cas simple, les adresses partiellement invalides sont :*

- « A_2R_2inv », le numéro est invalide dans la rue R_2 dans A_2 , et
- « A_2inv », la partie distribution est invalide dans A_2 .

Tous les singletons de Ω peuvent être représentés sur la figure 3.3. Les singletons de Ω sont notés identiquement au code de l'adresse correspondante. Les ensembles de singletons composant un acheminement (resp. une voie) sont notés par le code de l'acheminement (resp. de la voie avec son acheminement). Ainsi, $\Omega = A_2 \cup \{A_1\} \cup \{inv\}$, où $A_2 = A_2R_2 \cup \{A_2R_1, A_2B_1, A_2B_2, A_2inv\}$ et $A_2R_2 = \{A_2R_22, A_2R_24, A_2R_26, A_2R_2inv\}$.

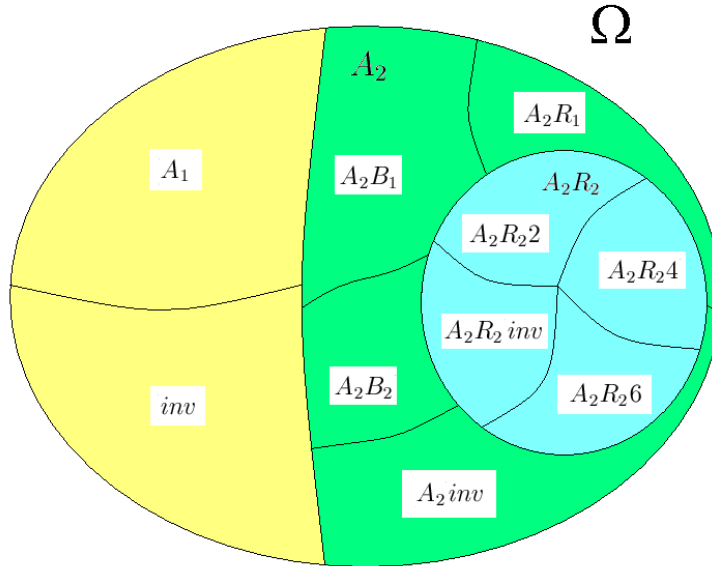


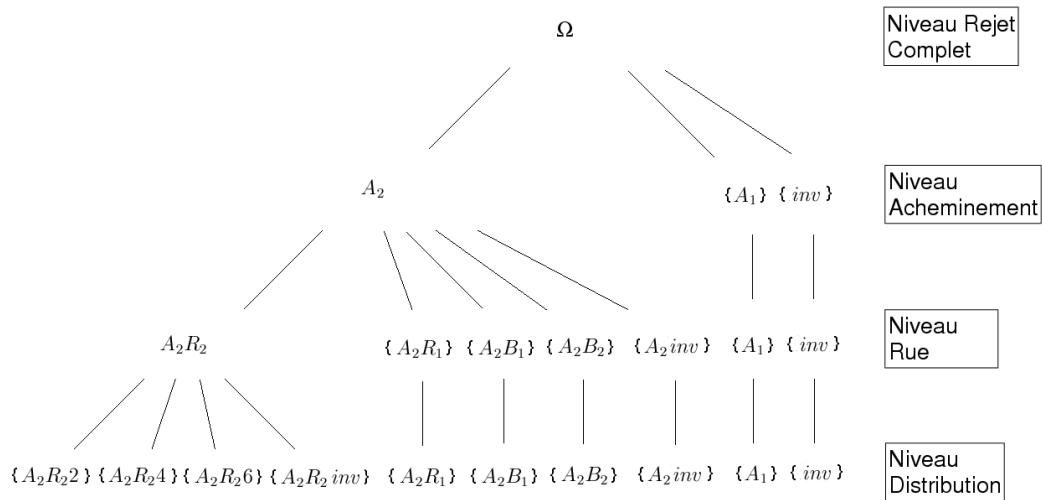
Figure 3.3 – Éléments du cadre de discernement associés à l'exemple 3.1.

L'ensemble \mathcal{D} des décisions susceptibles d'être prises par les LAP est composé :

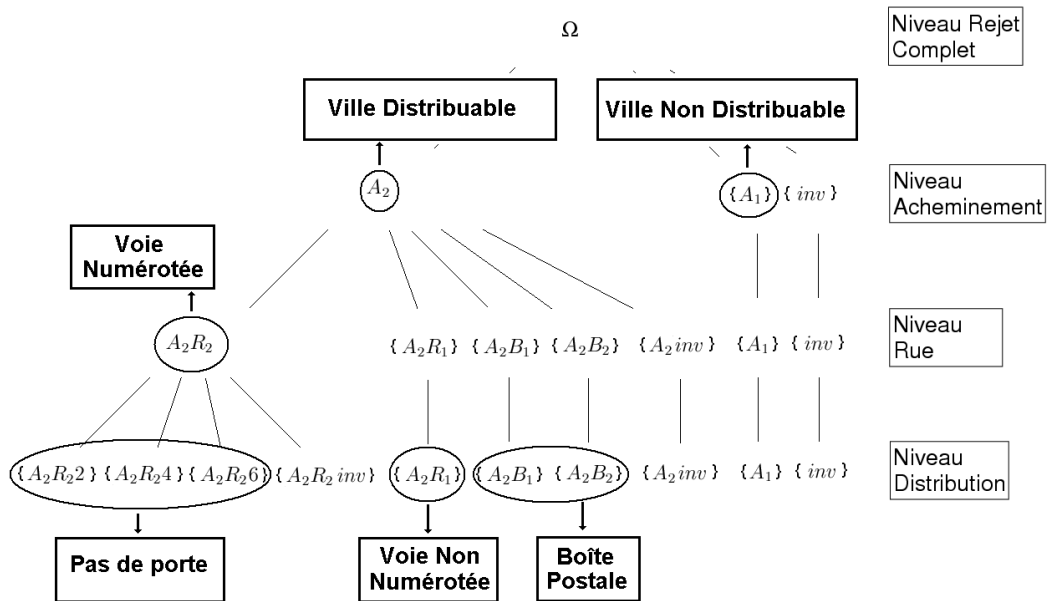
- des décisions en faveur d'adresses complètes différentes des adresses totalement ou partiellement invalides, qui sont des éléments singletons de Ω :
 - $\{A_1\} = \text{« } A_1 \text{ »}$,
 - $\{A_2B_1\} = \text{« } A_2B_1 \text{ »}$,
 - $\{A_2B_2\} = \text{« } A_2B_2 \text{ »}$,
 - $\{A_2R_1\} = \text{« } A_2R_1 \text{ »}$,
 - $\{A_2R_22\} = \text{« } A_2R_22 \text{ »}$,
 - $\{A_2R_24\} = \text{« } A_2R_24 \text{ »}$,
 - $\{A_2R_26\} = \text{« } A_2R_26 \text{ »}$, et,
- des décisions en faveur d'adresses partielles, qui sont des sous-ensembles de Ω :

- $\Omega = \text{« Rej »}$,
- $A_2 = \text{« } A_2\text{Rej »}$,
- $A_2R_2 = \text{« } A_2R_2\text{Rej »}$.

L'organisation hiérarchique de toutes les décisions peut être représentée sur la figure 3.4(a). La figure 3.4(b) illustre les liens entre cette hiérarchie et la hiérarchie de référence représentant uniquement les niveaux et les catégories de décisions.



(a) Hiérarchie associée à l'exemple 3.1 lié à une base de données réduite.



(b) Catégories des décisions des LAP dans l'exemple 3.1.

Figure 3.4 – Hiérarchie associée à l'exemple 3.1 et lien avec la hiérarchie de référence.

Toute information sur la véritable adresse sera modélisée par une fonction de croyance sur le cadre de discernement Ω . Par exemple, une information sur le fait que la véritable adresse se trouve dans une certaine ville sera modélisée par une

masse de croyance allouée à l'ensemble regroupant tous les singletons correspondant à des adresses situées dans cette ville.

Disposant de plusieurs LAP, le modèle de fusion proposé associe à la décision de chaque LAP une fonction de croyance sur Ω . Le paragraphe suivant expose la méthode de conversion de la sortie d'un LAP en fonction de croyance, à l'aide de la matrice de confusion de ce LAP.

3.2 Modèles hiérarchiques d'affectation

La forme purement hiérarchique de l'affectation développée dans ce cadre de décisions multi-niveaux est tout d'abord présentée. Une extension de ce modèle d'affectation où, à chaque niveau de la hiérarchie, des ensembles de classes peuvent être regroupés en catégorie, sera ensuite exposée.

3.2.1 Affectation hiérarchique

Définition 3.1 (Hiérarchie). *Une hiérarchie est un arbre vérifiant les conditions suivantes :*

- (i) *la racine est Ω ;*
- (ii) *l'ensemble des éléments de chaque niveau forme une partition de Ω ;*
- (iii) *chaque feuille est associée à un élément singleton.*

Par convention, le premier niveau est le niveau composé des singletons de Ω , et le dernier niveau est celui composé de Ω .

Notations :

- P désigne le nombre de niveaux de la hiérarchie ;
- $\Omega^{(p)}$, $p \in \{1, \dots, P\}$, désigne l'ensemble des décisions de niveau p de la hiérarchie ; par exemple, $\Omega^{(1)} = \Omega$ et $\Omega^{(P)} = \{\Omega\}$;
- $K^{(p)}$, $p \in \{1, \dots, P\}$, dénote le nombre d'éléments de $\Omega^{(p)}$;
- ω_k^p , $p \in \{1, \dots, P\}$, $k \in \{1, \dots, K^{(p)}\}$, désigne le $k^{\text{ième}}$ élément de $\Omega^{(p)}$.

Un exemple de hiérarchie à trois niveaux avec ces notations est illustré sur la figure 3.5.

Dans ce contexte de décisions multi-niveaux, il est possible de distinguer plusieurs types d'erreurs.

Exemple 3.2. *Supposons qu'un LAP fournisse une décision $\{A_2R_26\}$ située au niveau distribution de la hiérarchie représentée sur la figure 3.4(a).*

Si la vérité se trouve dans cette même rue mais à un numéro différent, par exemple $\{A_2R_22\}$, alors la décision est incorrecte. Cependant, la ville et la rue de la décision fournie sont communes avec la véritable adresse. Ainsi, nous dirons que cette décision, incorrecte au niveau distribution, est correcte au niveau rue. Dans la hiérarchie, le plus petit élément parent de la décision, contenant la vérité, est l'élément A_2R_2 de niveau rue.

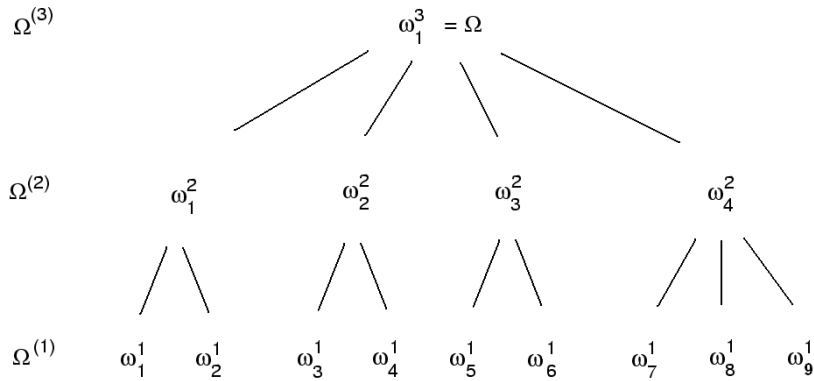


Figure 3.5 – Exemple d'une hiérarchie à trois niveaux.

Si cette même décision est fournie sur une lettre dont la vérité est $\{A_1\}$, le plus petit élément de la hiérarchie contenant cette décision et la vérité est Ω : cette décision, incorrecte au niveau distribution, n'est correcte qu'au niveau rejet complet. Aucun élément de l'adresse décidée n'est en commun avec la véritable adresse.

Si la décision avait été une décision de niveau acheminement A_2 , en faveur d'une adresse partielle « A_2Rej », au regard d'un envoi de vérité A_2R_2 . Alors cette décision aurait été correcte au niveau acheminement : la vérité et la décision sont dans le même élément de la hiérarchie au niveau acheminement.

Une décision de niveau p est ainsi dite correcte à un niveau $q \in [p, P]$ de la hiérarchie. Si la décision contient la vérité, le niveau q est égal à au niveau p ; sinon, q correspond au niveau du premier élément parent de la décision contenant la vérité. Une décision de niveau p correcte au niveau p , sera dite correcte. Nous noterons $n^{p,q}$ le nombre de décisions de niveau p correctes au niveau q où $q \in [p, P]$.

Exemple 3.3. *Considérons un classifieur fournissant des décisions en accord avec la hiérarchie illustrée sur la figure 3.5 et dont la matrice de confusion est représentée dans le tableau 3.1. Cette matrice de confusion contient des lignes en faveur de décisions de niveaux 1, 2 et 3. Les décisions de niveau 2 sont relatives à des ensembles de classes intermédiaires entre Ω et les singletons.*

Le nombre de décisions de niveau 1 correctes est égal à la somme des nombres inscrits en gras dans la matrice.

Le nombre de décisions de niveau 1 correctes au niveau 2 est égal à la somme des nombres inscrits en italique et soulignés dans la matrice. Ces nombres correspondent aux cas où la décision de niveau 1 fournie est incorrecte, mais la vérité se trouve dans l'élément parent dans la hiérarchie.

La somme des nombres restants à ce niveau correspond au nombre de décisions de niveau 1 correctes au niveau 3 : il faut remonter jusqu'à Ω pour que la décision et la vérité soient dans le même élément de la hiérarchie. En particulier, l'élément parent de la décision ne contient pas la vérité.

Ainsi :

$$\begin{aligned}
 n^{1,1} &= 34 + 23 + 24 + 28 + 35 + 15 + 20 + 35 + 31 = 245 , \\
 n^{1,2} &= 1 + 2 + 2 + 3 + 0 + 3 + 10 + 1 + 5 + 1 + 8 + 3 = 39 , \\
 n^{1,3} &= 1 \times 8 + 2 + 3 + 4 + 1 + 1 + 3 + 2 + 1 + 3 + 2 = 30 .
 \end{aligned} \tag{3.1}$$

Tableau 3.1 – Un exemple de matrice de confusion associée à un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 3.5.

Vérité	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9
Décisions									
Niveau 1									
$\omega_1^1 = \{\omega_1\}$	34	<u>1</u>	1	0	0	0	0	0	0
$\omega_2^1 = \{\omega_2\}$	<u>2</u>	23	0	0	0	0	1	0	0
$\omega_3^1 = \{\omega_3\}$	0	1	24	<u>2</u>	0	0	0	0	0
$\omega_4^1 = \{\omega_4\}$	1	0	<u>3</u>	28	0	0	0	0	0
$\omega_5^1 = \{\omega_5\}$	0	0	0	1	35	<u>0</u>	0	1	0
$\omega_6^1 = \{\omega_6\}$	0	0	0	1	<u>3</u>	15	0	0	1
$\omega_7^1 = \{\omega_7\}$	0	2	3	0	4	1	20	<u>10</u>	<u>1</u>
$\omega_8^1 = \{\omega_8\}$	0	1	0	0	3	2	<u>5</u>	35	<u>1</u>
$\omega_9^1 = \{\omega_9\}$	0	1	0	3	2	0	<u>8</u>	<u>3</u>	31
Niveau 2									
$\omega_1^2 = \{\omega_1, \omega_2\}$	10	12	0	0	1	0	1	0	0
$\omega_2^2 = \{\omega_3, \omega_4\}$	0	1	20	9	0	0	0	2	0
$\omega_3^2 = \{\omega_5, \omega_6\}$	0	0	0	0	12	24	0	0	0
$\omega_4^2 = \{\omega_7, \omega_8, \omega_9\}$	0	2	4	6	0	3	24	12	15
Niveau 3									
$\omega_1^3 = \Omega$	1	2	0	4	1	2	1	0	2

À partir de ces nombres, nous proposons une méthode d'affectation basée sur la distance entre la décision fournie par le classifieur et le plus petit élément de la hiérarchie contenant cette décision et la vérité.

Le nombre total d'objets classés au niveau p est noté par :

$$n^p = \sum_{q=p}^P n^{p,q} . \tag{3.2}$$

Soit u la fonction qui, à un élément de la hiérarchie différent de Ω , associe son élément parent situé juste au dessus dans la hiérarchie. Par exemple, avec la hiérarchie représentée sur la figure 3.5, $u(\omega_3^1) = \omega_2^2$, et $u(\omega_2^2) = \omega_1^3$.

Sachant que le classifieur fournit une décision d de niveau p , l'**affectation hiérarchique** proposée initialise m par :

$$\begin{aligned}
 m : \quad 2^\Omega &\longrightarrow [0, 1] \\
 u^{q-p}(d) &\longmapsto \frac{n^{p,q}}{n^p} \quad \forall q \in [p, P] ,
 \end{aligned} \tag{3.3}$$

où $u^{q-p} = \underbrace{u \circ \dots \circ u \circ u}_{q-p \text{ fois}}$, et u^0 est la fonction identité.

Notons que si $d = \omega_1^P = \Omega$, alors $m(d) = m(\Omega) = \frac{n^{P,P}}{n^P} = \frac{n^P}{n^P} = 1$.

Les fonctions de masse générées par cette affectation sont consonantes.

Exemple 3.4 (suite de l'exemple 3.3). *Supposons que le classifieur de l'exemple 3.3 fournisse la décision, de niveau 1, $\omega_3^1 = \{\omega_3\}$. Le nombre total de décisions de niveau 1 fournies est $n^1 = n^{1,1} + n^{1,2} + n^{1,3} = 314$. L'affectation (3.3) conduit à représenter la décision de ce classifieur par la fonction de masse m , telle que :*

$$\begin{aligned} m(\omega_3^1) &= m(\{\omega_3\}) &= n^{1,1}/n^1 &= 245/314 &= 0.78 , \\ m(u(\omega_3^1)) &= m(\{\omega_3, \omega_4\}) &= n^{1,2}/n^1 &= 39/314 &= 0.12 , \\ m(u^2(\omega_3^1)) &= m(\Omega) &= n^{1,3}/n^1 &= 30/314 &= 0.10 . \end{aligned} \quad (3.4)$$

3.2.2 Affectation hiérarchique catégorisée

Par l'affectation précédente, les décisions d'un classifieur attachées à un même niveau de la hiérarchie sont représentées par une fonction de masse avec des poids de même valeur. Or, toutes les décisions d'un même niveau n'ont pas la même complexité. Par exemple, au niveau distribution de la hiérarchie représentée sur la figure 3.4(a), la détection d'une voie numérotée est plus complexe que la détection d'une voie non numérotée où le numéro de pas de porte n'est pas à identifier. De même, les décisions en faveur d'adresses comportant une distribution spéciale, où les intitulés sont moins nombreux et généralement plus étudiés que les intitulés de voie, possèdent des performances différentes de celles en faveur d'adresses en faveur de voies numérotées ou non.

Afin de prendre en compte ces différences de réussite, des *catégories* de décisions sont introduites et représentées dans la hiérarchie. L'affectation précédente est alors raffinée en se basant sur le niveau de la décision proposée par le classifieur, et sur la catégorie à ce niveau de cette décision.

Dans ce paragraphe, la notion de hiérarchie est étendue. À chaque niveau de la hiérarchie, des groupes peuvent être distingués, définissant une *catégorie* ou famille d'ensembles de classes et vérifiant une règle : les fils de deux éléments de catégories différentes ne peuvent pas être de même catégorie ; la représentation de la hiérarchie en considérant uniquement les niveaux et les catégories des éléments est donc aussi un arbre. La hiérarchie de référence illustrée sur la figure 3.2 constitue cet arbre des catégories et niveaux considérés dans cette application postale.

Chaque décision est un élément situé à un niveau de la hiérarchie, et appartient à une certaine catégorie. La hiérarchie représentée sur la figure 3.6 illustre une hiérarchie à trois niveaux avec des catégories de décisions.

L'affectation hiérarchique catégorisée se base sur la même idée de distance que l'affectation précédente. Les mêmes éléments focaux seront créés. Seules les valeurs des masses seront dépendantes du niveau et de la catégorie de la décision fournie par le classifieur.

Soit $n_c^{p,q}$ le nombre de décisions de niveau p et de catégorie c correctes au niveau q où $q \in [p, P]$.

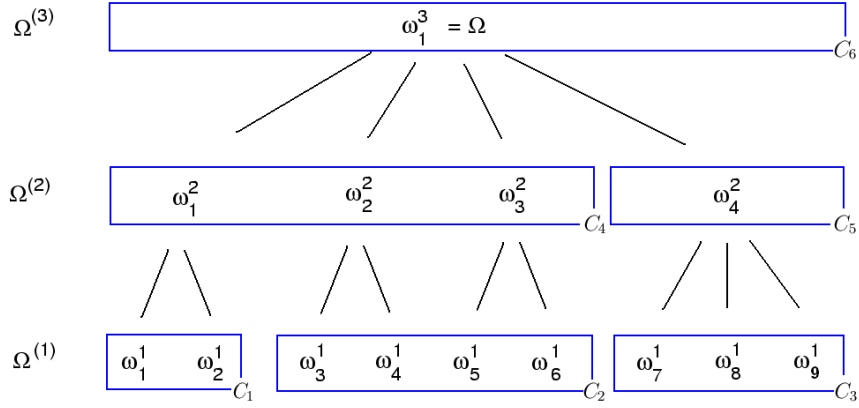


Figure 3.6 – Un exemple de hiérarchie à trois niveaux avec des catégories de décisions.

Le nombre total d'objets classés au niveau p de catégorie c est noté :

$$n_c^p = \sum_{q=p}^P n_c^{p,q} . \quad (3.5)$$

Sachant que le classifieur fournit une décision d de niveau p et de catégorie c , l'affectation hiérarchique catégorisée proposée, initialise m par :

$$m : \begin{array}{ll} 2^\Omega & \longrightarrow [0, 1] \\ u^{q-p}(d) & \longmapsto \frac{n_c^{p,q}}{n_c} \quad \forall q \in [p, P] . \end{array} \quad (3.6)$$

Notons que si $d = \omega_1^P = \Omega$, comme il n'est possible de définir qu'une seule catégorie au dernier niveau de la hiérarchie : $m(d) = m(\Omega) = \frac{n_c^{P,P}}{n_c} = \frac{n^{P,P}}{n^P} = \frac{n^P}{n^P} = 1$.

Exemple 3.5. *Considérons un classifieur fournissant des décisions en accord avec la hiérarchie de la figure 3.6 et dont la matrice de confusion, fournie dans le tableau 3.2, comporterait les mêmes valeurs que dans l'exemple 3.4 (tableau 3.1). La matrice de confusion contient toujours des lignes en faveur de décisions de niveaux 1, 2 et 3. Mais à chaque niveau, des catégories de décisions ont été différenciées.*

Supposons que le classifieur décide $\omega_3^1 = \{\omega_3\}$. Cette décision est de niveau 1 et de catégorie c_2 . Le nombre de décisions de niveau 1 et de catégorie c_2 correctes est égal à la somme des nombres inscrits en gras dans la matrice (tableau 3.2). Le nombre de décisions de niveau 1 et de catégorie c_2 correctes au niveau 2 est égal à la somme des nombres inscrits en italique et soulignés, et la somme des nombres situés sur ces mêmes 4 lignes et distincts de ceux précédemment énoncés est égale au nombre de décisions de niveau 1 et de catégorie c_2 correctes au niveau 3.

$$\begin{aligned} n_{c_2}^{1,1} &= 24 + 28 + 35 + 15 &= 102 , \\ n_{c_2}^{1,2} &= 2 + 3 + 0 + 3 &= 8 , \\ n_{c_2}^{1,3} &= 1 + 1 + 1 + 1 + 1 + 1 &= 6 , \end{aligned} \quad (3.7)$$

Tableau 3.2 – Matrice de confusion d'un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 3.6.

Vérité	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9
Décisions									
Niveau 1									
catégorie c_1									
$\omega_1^1 = \{\omega_1\}$	34	1	1	0	0	0	0	0	0
$\omega_2^1 = \{\omega_2\}$	2	23	0	0	0	0	1	0	0
catégorie c_2									
$\omega_3^1 = \{\omega_3\}$	0	1	24	<u>2</u>	0	0	0	0	0
$\omega_4^1 = \{\omega_4\}$	1	0	<u>3</u>	28	0	0	0	0	0
$\omega_5^1 = \{\omega_5\}$	0	0	0	1	35	<u>0</u>	0	1	0
$\omega_6^1 = \{\omega_6\}$	0	0	0	1	<u>2</u>	15	0	0	1
catégorie c_3									
$\omega_7^1 = \{\omega_7\}$	0	2	3	0	4	1	20	10	1
$\omega_8^1 = \{\omega_8\}$	0	1	0	0	3	2	5	35	1
$\omega_9^1 = \{\omega_9\}$	0	1	0	3	2	0	8	3	31
Niveau 2									
catégorie c_4									
$\omega_1^2 = \{\omega_1, \omega_2\}$	10	12	0	0	1	0	1	0	0
$\omega_2^2 = \{\omega_3, \omega_4\}$	0	1	20	9	0	0	0	2	0
$\omega_3^2 = \{\omega_5, \omega_6\}$	0	0	0	0	12	24	0	0	0
catégorie c_5									
$\omega_4^2 = \{\omega_7, \omega_8, \omega_9\}$	0	2	4	6	0	3	24	12	15
Niveau 3									
catégorie c_6									
$\omega_1^3 = \Omega$	1	2	0	4	1	2	1	0	2

et, la fonction de masse attribuée à ce classifieur sera :

$$\begin{aligned} m(\omega_3^1) &= m(\{\omega_3\}) &= n_{c_2}^{1,1}/n_{c_2}^1 &= 102/116 = 0.88 , \\ m(u(\omega_3^1)) &= m(\{\omega_3, \omega_4\}) &= n_{c_2}^{1,2}/n_{c_2}^1 &= 8/116 = 0.07 , \\ m(u^2(\omega_3^1)) &= m(\Omega) &= n_{c_2}^{1,3}/n_{c_2}^1 &= 6/116 = 0.05 . \end{aligned} \quad (3.8)$$

Par rapport au résultat obtenu avec l'affectation hiérarchique précédente (équation 3.4), basée uniquement sur le niveau des décisions, la masse allouée à $\{\omega_3\}$ est plus importante. La classe ω_3 appartient à un sous-ensemble de décisions de niveau 1 plus fiable que l'ensemble des décisions de niveau 1 de cette hiérarchie. Par exemple, les décisions de niveau 1 et de catégorie c_3 sont moins fiables : si ce classifieur décide une décision de niveau 1 et de catégorie c_3 , par exemple $\omega_9^1 = \{\omega_9\}$, alors la fonction de masse attribuée à ce classifieur pour cette décision sera :

$$\begin{aligned} m(\omega_9^1) &= m(\{\omega_9\}) &= n_{c_3}^{1,1}/n_{c_3}^1 &= 86/136 = 0.63 , \\ m(u(\omega_9^1)) &= m(\{\omega_7, \omega_8, \omega_9\}) &= n_{c_3}^{1,2}/n_{c_3}^1 &= 28/136 = 0.21 , \\ m(u^2(\omega_9^1)) &= m(\Omega) &= n_{c_3}^{1,3}/n_{c_3}^1 &= 22/136 = 0.16 . \end{aligned} \quad (3.9)$$

L'affectation proposée est intermédiaire entre une affectation Bayésienne offrant à tous les singletons de Ω la possibilité d'une masse positive, mais nécessitant un ensemble d'apprentissage de taille élevé, et l'affectation de Xu et al, attribuant à toutes les décisions différentes de Ω la même masse.

Remarque 3.1.

1. Si la hiérarchie ne comporte qu'une seule catégorie par niveau, l'affectation catégorisée (3.6) est équivalente à l'affectation hiérarchique (3.3).
2. Si la hiérarchie ne comporte que deux niveaux avec un premier niveau composé de tous les singletons de Ω tous d'une même catégorie, alors l'affectation catégorisée (3.6) est équivalente à l'affectation basée sur le taux de fiabilité (2.12).
3. Si la hiérarchie ne comporte que deux niveaux avec un premier niveau des singletons de Ω , tous attachés à une catégorie propre, alors l'affectation développée (3.6) coïncide avec l'affectation Bayésienne sur la décision du classifieur mais diffère sur les autres éléments.
4. L'affectation (3.6) fournit toujours une masse consonante.

Remarque 3.2. Cette méthode d'affectation peut être employée sans qu'une hiérarchie de décisions soit explicite. La définition d'une hiérarchie sur un problème où un classifieur ne fournirait des décisions qu'aux premier et dernier niveau, permettrait d'allouer de la masse à des éléments intermédiaires entre les singletons et Ω . Une hiérarchie peut par exemple être définie sur l'exemple 2.3.2 concernant un classifieur en charge de la reconnaissance de caractères numériques. En supposant que certains chiffres sont souvent confondus, par exemple le chiffre 9 et le chiffre 4. Ceux-ci peuvent être virtuellement regroupés dans un ensemble de classes $\{4, 9\}$ à un niveau intermédiaire de la hiérarchie. À partir de la même matrice de confusion représentée par le tableau 2.3, et en définissant une catégorie ne regroupant que ces

deux chiffres, l'affectation basée sur cette hiérarchie à trois niveaux entraîne, dans le cas d'une décision en faveur du chiffre $\{9\}$, $m(\{9\}) = 13/19$, $m(\{4, 9\}) = 4/19$ et $m(\Omega) = 2/19$. Cette affectation des masses est intermédiaire entre les résultats de l'affectation Bayésienne (exemple 2.3.2) menant à $m_i(\{3\}) = 2/15$, $m_i(\{4\}) = 4/15$, $m_i(\{9\}) = 9/15$, et les résultats obtenus par l'affectation de Xu et al. (exemple 2.3) menant à $m_i(\{9\}) = 60/96$, $m_i(\Omega \setminus \{9\}) = 20/96$, $m_i(\Omega) = 16/96$.

Pour s'adapter aux particularités de l'application postale développée, des ajustements de cette affectation ont été réalisés.

3.2.3 Affectation postale employée

Dans notre application, les LAP ne décident pas les adresses totalement ou partiellement invalides. Ainsi, l'emploi de l'affectation précédente impliquerait que ces classes ne recevraient jamais de poids, or ces classes existent et jouent un rôle important. Une lettre comportant une adresse invalide ne peut être comptée qu'en rejet ou en erreur. Si des informations sont disponibles sur l'invalidité d'une lettre, nous voulons les prendre en compte pour ne pas favoriser une décision en faveur d'une adresse valide. Ainsi, l'affectation précédente est quelque peu modifiée de façon à ce que les décisions relatives à des adresses invalides puissent recevoir de la masse.

Définition 3.2 (Niveau d'une adresse invalide). *Le niveau d'une adresse invalide est défini comme le niveau du plus petit élément parent dans la hiérarchie qui ne soit pas relatif à une adresse invalide, moins 1.*

Exemple 3.6. *En reprenant l'exemple 3.1, par la hiérarchie illustrée sur la figure 3.4(b) :*

- « A_2R_2inv » est une adresse invalide de niveau 1,
- « A_2inv » est une adresse invalide de niveau 2,
- « inv » est une adresse invalide de niveau 3.

Soit $n_c^{p,inv(q)}$ le nombre de fois où la décision fournie était de niveau p et de catégorie c , et la vérité était invalide de niveau q .

Sachant que le classifieur fournit une décision d de niveau p et de catégorie c , l'**affectation postale** employée, initialise m par :

$$\begin{aligned}
 m : \quad 2^\Omega &\longrightarrow [0, 1] \\
 u^{q-p}(d) &\longmapsto \frac{n_c^{p,q} - n_c^{p,inv(q-1)}}{n_c^p} \quad \forall q \in [p, P], \\
 u^{q-p}(d)inv &\longmapsto \frac{n_c^{p,inv(q-1)}}{n_c^p} \quad \forall q \in [p, P],
 \end{aligned} \tag{3.10}$$

où, par convention, $n_c^{p,inv(0)} = 0$, et « Ωinv » = « inv ».

Pour tout $A \subseteq \Omega$ tel que $|A| > 1$, nous dirons que « $A inv$ » est l'invalide associé à « A ».

Exemple 3.7. *Considérons la matrice de confusion illustrée dans le tableau 3.3, associée à un LAP de l'exemple 3.1. Soit d la décision fournie par ce LAP au regard d'un envoi à identifier.*

- Si $d = \Omega$. Cette décision est de niveau 4, et appartient à l'unique catégorie de ce dernier niveau de la hiérarchie. L'information issue du classifieur sera représentée par la fonction de masses m définie par :

$$\begin{cases} m(\{inv\}) &= \frac{n^{4,inv(3)}}{n^4} &= \frac{25}{100} &= 0.25 , \\ m(\Omega) &= \frac{n^{4,4} - n^{4,inv(3)}}{n^4} &= \frac{(25+12+10+8+3+7+5+20+9+1)-25}{100} &= 0.75 . \end{cases}$$

Tableau 3.3 – Matrice de confusion associée à un LAP sur la base de données réduite de l'exemple 3.1. Notations : cat = catégorie, bp = boîtes postales, vnn = voies non numérotées, pp = pas de portes, vn = voies numérotées, ad = acheminement distribuable, and = acheminement non distribuable.

Niveau	Cat	Vérité Décision	<i>inv</i>	A_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2
			<i>inv</i>	B_1	B_2	R_1	R_2	R_2	R_2	R_2	R_2	
								2	4	6	<i>inv</i>	
$(^1)\Omega$ (Distri- bution)	bp	A_2B_1	0	0	1	95	3	0	0	0	1	0
		A_2B_2	1	0	0	2	96	0	0	0	0	1
	vnn	A_2R_1	0	1	0	0	0	97	1	1	0	0
		A_2R_22	0	0	0	3	0	0	90	1	5	1
		A_2R_24	0	2	0	0	0	3	6	88	1	0
A_2R_26	0	4	1	0	1	2	7	1	83	1		
$(^2)\Omega$ (Rue)	vn	A_2R_2	0	1	0	0	0	1	17	31	42	8
$(^3)\Omega$ (Achemi- nement)	and	A_1	0	90	1	1	0	1	2	2	2	1
	ad	A_2	0	1	4	5	5	40	32	8	4	1
$(^4)\Omega$ (Rejet Complet)		Ω	25	12	10	8	3	7	5	20	9	1

- Si $d = \{A_2R_22\}$. Cette décision est de niveau 1 et de catégorie $c = \ll \text{pas de porte} \gg$. Ainsi :

$$\begin{cases} m(\{A_2R_22\}) &= \frac{n_c^{1,1}}{n_c^1} &= \frac{90+88+83}{300} &= 0.870 , \\ m(\{A_2R_2inv\}) &= \frac{n_c^{1,inv(1)}}{n_c^1} &= \frac{1+1}{300} &= 0.007 , \\ m(A_2R_2) &= \frac{n_c^{1,2} - n_c^{1,inv(1)}}{n_c^1} &= \frac{(1+5+1+6+1+7+1+1)-(1+1)}{300} &= 0.070 , \\ m(\{A_2inv\}) &= \frac{n_c^{1,inv(2)}}{n_c^1} &= \frac{1}{300} &= 0.003 , \\ m(A_2) &= \frac{n_c^{1,3} - n_c^{1,inv(2)}}{n_c^1} &= \frac{(3+3+1+1+2)-1}{300} &= 0.030 , \\ m(\{inv\}) &= \frac{n_c^{1,inv(3)}}{n_c^1} &= \frac{0}{300} &= 0.000 , \\ m(\Omega) &= \frac{n_c^{1,4} - n_c^{1,inv(3)}}{n_c^1} &= \frac{2+4}{300} &= 0.020 . \end{cases}$$

Avec cette affectation, même un rejet complet peut apporter une information différente de la fonction de masse vide, une information sur la validité de la lettre est gardée. En général, quand un LAP effectue un rejet complet, il y a plus de raisons de penser que la lettre est invalide, ou du moins que la lettre est difficile à lire, que si un LAP fournit une décision avec des éléments d’une adresse. Le degré de vérité de cette observation se retrouve dans la matrice de confusion du LAP, et ainsi dans cette conversion de la décision du LAP en une fonction de croyance.

Cette affectation ne produit généralement plus de masses consonantes, cependant les masses allouées à des éléments relatifs à des adresses invalides intermédiaires entre la décision et Ω sont généralement très faibles. En réalité, cette affectation trouve principalement son utilité lorsque les décisions fournies sont partielles. Dans ces cas, les masses allouées aux invalides sont significatives. Cette affectation traduit l’information du rejet en une part de croyance sur le fait « rejet car invalide » et une part de croyance sur le rejet en ambiguïté.

Enfin, comme une indication sur le type d’écriture de l’envoi est disponible, deux matrices de confusion sont créées pour chaque LAP : une pour le courrier dactylographié, et une pour le courrier manuscrit. Lorsque le LAP fournit une décision, les données utilisées pour l’affectation de la masse seront celles du type d’écriture indiqué en plus de cette décision.

3.2.4 Combinaison des fonctions de croyance

Dans ce modèle de base, les LAP sont supposés distincts, et l’affectation fiable, ainsi la règle de combinaison conjonctive est appliquée.

Exemple 3.8 (Combinaison de 3 LAP). *Considérons 3 LAP, notés 1, 2, et 3, ayant formulé les décisions d_1, d_2 , et d_3 suivantes au regard d’une image d’un même envoi postal :*

- $d_1 = A_2$, i.e., une décision de niveau acheminement et de catégorie « ville distribuable » ;
- $d_2 = \{A_1\}$, i.e., une décision de niveau acheminement et de catégorie « ville non distribuable » ;
- $d_3 = \{A_2R_117\}$, i.e., une décision de niveau distribution et de catégorie « pas de porte ».

En notant m_i la masse issue d’un LAP i , pour tout $i \in \{1, 2, 3\}$, et en supposant que les principes d’affectation des masses présentés précédemment conduisent à :

$$\begin{cases} m_1(A_2) & = 0.95 \\ m_1(\{A_2inv\}) & = 0.04 \\ m_1(\Omega) & = 0.01 \end{cases} \quad \begin{cases} m_2(\{A_1\}) & = 0.9 \\ m_2(\Omega) & = 0.1 \end{cases}$$

$$\begin{cases} m_3(\{A_2R_117\}) & = 0.88 \\ m_3(A_2R_1) & = 0.07 \\ m_3(A_2) & = 0.03 \\ m_3(\Omega) & = 0.02 \end{cases},$$

le calcul de $m = m_1 \odot m_2 \odot m_3$ donne :

$$\begin{aligned}
 m(\{A_2R_117\}) &= 0.0845 & m(\{A_1\}) &= 0.0002 \\
 m(A_2R_1) &= 0.0067 & m(\Omega) &= 0.0000 \\
 m(\{A_2inv\}) &= 0.0002 & m(\emptyset) &= 0.9036 . \\
 m(A_2) &= 0.0048 & &
 \end{aligned} \tag{3.11}$$

La masse importante portée par le vide reflète le conflit important entre les décisions fournies.

3.3 Prise de décision

L'ensemble \mathcal{D} des décisions pouvant être prises par la combinaison est le même que celui des LAP (cf exemple 3.1). L'ensemble \mathcal{D} est composée des décisions en faveur d'adresses complètes différentes des adresses totalement ou partiellement invalides, et des décisions en faveur d'adresses partielles. Les décisions peuvent être de niveau distribution, rue, acheminement ou rejet complet.

Du fait de la grande cardinalité de la base de données et de sa complexité, nous ne pouvons connaître précisément la cardinalité de chaque sous-ensemble de Ω en un temps raisonnable, ce qui nous empêche de calculer la probabilité pignistique sur Ω attachée au risque espéré de chaque décision (équation 2.32). Toutefois, nous pouvons prendre une décision sur un cadre de pari Γ calculé à partir des éléments focaux des masses des LAP et de la hiérarchie.

Exemple 3.9 (suite de l'exemple 3.8). *Au vu des décisions des LAP, les décisions que la combinaison s'autorise à prendre sont :*

- $d_1 = \Omega$ en faveur du code « *Rej* »,
- $d_2 = \{A_1\}$ en faveur du code « *A₁* »,
- $d_3 = A_2$ en faveur du code « *A₂Rej* »,
- $d_4 = A_2R_1$ en faveur du code « *A₂R₁Rej* »,
- et $d_5 = A_2R_117$ en faveur du code « *A₂R₁17* ».

Le cadre de pari Γ associé à cet envoi est représenté sur la figure 3.7. Il est composé des éléments :

- $\gamma_1 = \{A_2R_117\}$,
- $\gamma_2 = \{A_2R_1inv\}$,
- $\gamma_3 = A_2R_1 \setminus \{A_2R_117, A_2R_1inv\}$,
- $\gamma_4 = \{A_2inv\}$,
- $\gamma_5 = A_2 \setminus (A_2R_1 \cup \{A_2inv\})$,
- $\gamma_6 = \{A_1\}$,
- $\gamma_7 = \{inv\}$
- et $\gamma_8 = \Omega \setminus (A_2 \cup \{A_1, inv\})$.

La construction du cadre de pari est décrite dans l'algorithme 1. L'exemple 3.9 est détaillé en annexe ??.

Outre le coût de bonne réponse, ici égal à zéro, les coûts de décision peuvent se regrouper en coûts de rejet et coûts d'erreur de niveaux acheminement, rue et distribution, définis de la manière suivante.

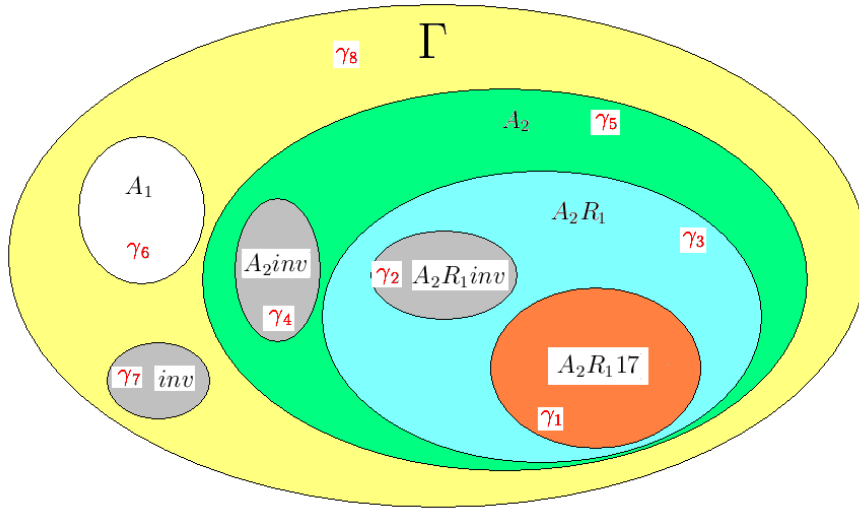


Figure 3.7 – Éléments du cadre de pari Γ construit à partir de Ω , de la hiérarchie de référence (figure 3.2), et des masses associées à l'exemple 3.8.

Algorithme 1 : Construction du cadre de pari Γ .

Données :

m : une fonction de masse;

H : hiérarchie;

Γ : cadre de pari à construire;

// variables

$\mathcal{F}(m)$: un ensemble contenant les éléments focaux de m ;

γ : un sous-ensemble de Ω , un élément singleton de Γ ;

F : un sous-ensemble de Ω ;

// instructions

ajouter à $\mathcal{F}(m)$:

– Ω , s'il n'est pas déjà présent ;

– tous les invalides non déjà présents associés aux éléments de $\mathcal{F}(m)$;

$\Gamma \leftarrow \emptyset$;

répéter

$F \leftarrow$ un élément de $\mathcal{F}(m)$ de niveau le plus élevé dans H ;

$\gamma \leftarrow F \setminus \{A \in \mathcal{F}(m) \text{ t.q. } F \text{ est le plus petit contenant de } A \text{ dans } \mathcal{F}(m)\}$;

 ajouter γ à Γ ;

 enlever F de $\mathcal{F}(m)$;

jusqu'à $\mathcal{F}(m) = \emptyset$;

- Un *coût de rejet* acheminement $\mathcal{C}_{R_{ache}}$ (resp. rue $\mathcal{C}_{R_{rue}}$, distribution $\mathcal{C}_{R_{dis}}$) correspond au prix à payer pour avoir fourni une décision en faveur d'une adresse partielle, de niveau rejet complet (resp. acheminement, rue) correcte.
- Un *coût d'erreur* de niveau acheminement $\mathcal{C}_{E_{ach}}$, (resp. rue $\mathcal{C}_{E_{rue}}$, distribution $\mathcal{C}_{E_{dis}}$) correspond au prix à payer pour avoir fourni une décision de niveau acheminement (resp. rue, distribution), correcte au niveau rejet complet (resp. acheminement, rue).

Exemple 3.10 (suite de l'exemple 3.9). *Les coûts associés à cet exemple sont indiqués dans le tableau 3.4.*

Tableau 3.4 – Coûts de décisions en fonction de la décision fournie et de la vérité.

Vérité	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
Décisions								
$d_1 = \langle \text{Rej} \rangle$	$\mathcal{C}_{R_{ach}}$	$\mathcal{C}_{R_{ach}}$	$\mathcal{C}_{R_{ach}}$	$\mathcal{C}_{R_{ach}}$	$\mathcal{C}_{R_{ach}}$	$\mathcal{C}_{R_{ach}}$	0	0
$d_2 = \langle A_1 \rangle$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	0	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$
$d_3 = \langle A_2 \text{Rej} \rangle$	$\mathcal{C}_{R_{rue}}$	$\mathcal{C}_{R_{rue}}$	$\mathcal{C}_{R_{rue}}$	0	0	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$
$d_4 = \langle A_2 R_1 \text{Rej} \rangle$	$\mathcal{C}_{R_{dis}}$	0	0	$\mathcal{C}_{E_{rue}}$	$\mathcal{C}_{E_{rue}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$
$d_5 = \langle A_2 R_1 17 \rangle$	0	$\mathcal{C}_{E_{dis}}$	$\mathcal{C}_{E_{dis}}$	$\mathcal{C}_{E_{rue}}$	$\mathcal{C}_{E_{rue}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$	$\mathcal{C}_{E_{ach}}$

La relation d'ordre suivante, conforme au sens commun, est définie :

$$0 \leq \mathcal{C}_{R_{dis}} \leq \mathcal{C}_{R_{rue}} \leq \mathcal{C}_{R_{ach}} \leq \mathcal{C}_{E_{dis}} \leq \mathcal{C}_{E_{rue}} \leq \mathcal{C}_{E_{ach}}. \quad (3.12)$$

Par exemple, il est moins grave de commettre une erreur sur la partie distribution que sur la ville, et mieux vaut rejeter la ville plutôt que de faire une erreur sur cette ville. Bien sûr, cet ordre dépend de l'application. Par exemple, il est possible que le coût du rejet de la ville soit plus élevé que le coût d'une erreur distribution. En effet, une erreur portant sur le point de distribution, où le lieu d'acheminement est correct, peut avoir un coût moins élevé que le travail demandé par un rejet global.

Exemple 3.11 (suite de l'exemple 3.10). *La probabilité pignistique, calculée sur Γ à partir de m (équation 3.11), est donnée par :*

$$\begin{aligned}
BetP(\{\gamma_1\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + \frac{m(A_2 R_1)}{3} + m(\{A_2 R_1 17\})\right) = 0.910 \\
BetP(\{\gamma_2\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + \frac{m(A_2 R_1)}{3} + m(\{A_2 R_1 inv\})\right) = 0.033 \\
BetP(\{\gamma_3\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + \frac{m(A_2 R_1)}{3}\right) = 0.033 \\
BetP(\{\gamma_4\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + m(\{A_2 inv\})\right) = 0.012 \\
BetP(\{\gamma_5\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + \frac{m(A_2)}{5}\right) = 0.010 \\
BetP(\{\gamma_6\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + m(\{A_1\})\right) = 0.002 \\
BetP(\{\gamma_7\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8} + m(\{inv\})\right) = 0.000 \\
BetP(\{\gamma_8\}) &= \mathcal{K}\left(\frac{m(\Omega)}{8}\right) = 0.000
\end{aligned} \quad (3.13)$$

avec $\mathcal{K} = \frac{1}{1-m(\emptyset)}$.

En posant $I = \{1, \dots, 8\}$, les expressions des risques sont alors les suivantes :

$$\begin{aligned}\rho_{bet}(d_1) &= \sum_{\gamma \in \Gamma} c(d_1, \gamma) BetP(\gamma) = \sum_{i \in \Gamma \setminus \{7,8\}} \mathcal{C}_{R_{ach}} BetP(\{\gamma_i\}) \\ &= \mathcal{C}_{R_{ach}} ,\end{aligned}$$

$$\begin{aligned}\rho_{bet}(d_2) &= \sum_{\gamma \in \Gamma} c(d_2, \gamma) BetP(\gamma) = \sum_{i \in \Gamma \setminus \{6\}} \mathcal{C}_{E_{ach}} BetP(\{\gamma_i\}) \\ &= 0.998 \mathcal{C}_{E_{ach}} ,\end{aligned}$$

$$\begin{aligned}\rho_{bet}(d_3) &= \sum_{i \in \{1,2,3\}} \mathcal{C}_{R_{rue}} BetP(\{\gamma_i\}) + \sum_{i \in \{6,7,8\}} \mathcal{C}_{E_{ach}} BetP(\{\gamma_i\}) \\ &= 0.976 \mathcal{C}_{R_{rue}} + 0.002 \mathcal{C}_{E_{ach}} ,\end{aligned}$$

$$\begin{aligned}\rho_{bet}(d_4) &= \mathcal{C}_{R_{dis}} BetP(\{\gamma_1\}) + \sum_{i \in \{4,5\}} \mathcal{C}_{E_{rue}} BetP(\{\gamma_i\}) \\ &\quad + \sum_{i \in \{6,7,8\}} \mathcal{C}_{E_{ach}} BetP(\{\gamma_i\}) \\ &= 0.910 \mathcal{C}_{R_{dis}} + 0.022 \mathcal{C}_{E_{rue}} + 0.002 \mathcal{C}_{E_{ach}} ,\end{aligned}$$

$$\begin{aligned}\rho_{bet}(d_5) &= \sum_{i \in \{2,3\}} \mathcal{C}_{E_{dis}} BetP(\{\gamma_i\}) + \sum_{i \in \{4,5\}} \mathcal{C}_{E_{rue}} BetP(\{\gamma_i\}) \\ &\quad + \sum_{i \in \{6,7,8\}} \mathcal{C}_{E_{ach}} BetP(\{\gamma_i\}) \\ &= 0.066 \mathcal{C}_{E_{dis}} + 0.022 \mathcal{C}_{E_{rue}} + 0.002 \mathcal{C}_{E_{ach}} .\end{aligned}$$

Si les coûts d'erreur ne sont pas trop élevés par rapport aux coûts de rejet, la décision sera d_5 du fait de l'ordre entre les coûts (3.12). Sinon la décision peut être d_1 , d_3 ou d_4 . Plus les coûts d'erreur seront élevés par rapport aux coûts de rejet, plus la combinaison aura tendance à rejeter. Par exemple, avec le 6-uplet $(\mathcal{C}_{R_{ach}}, \mathcal{C}_{E_{dis}}, \mathcal{C}_{E_{rue}}, \mathcal{C}_{E_{ach}})$ égal à :

- $(1, 2, 3, 4, 5, 6)$, la décision choisie est d_5 i.e. « A_2R_117 » ;
- $(1, 2, 3, 20, 25, 50)$, la décision choisie est d_4 i.e. « A_2R_1Rej » ;
- $(1, 2, 3, 40, 50, 60)$, la décision choisie est d_3 i.e. « A_2Rej » ;
- $(1, 2, 3, 400, 500, 600)$, la décision choisie est d_1 i.e. « Rej ».

Idéalement ces coûts sont donnés par des experts du domaine ou le client, et reflètent des coûts financiers réels. Mais si le client préfère une combinaison avec un taux d'erreur contrôlé à certains niveaux, ces coûts peuvent être déterminés à partir d'un ensemble d'apprentissage de manière à obtenir un comportement attendu de la combinaison.

3.4 Résultats expérimentaux

Dans ce mémoire, nous illustrons notre démarche de combinaison en considérant trois LAP, nommés LAP 1, LAP 2 et LAP 3. Les LAP 1 et 2 sont d'origine Solystic. Le LAP 3 est d'origine extérieure.

3.4.1 Choix des coûts

Les coûts sont appris manuellement de façon à obtenir le meilleur taux de lecture possible en maintenant le taux d'erreur inférieur dans un intervalle jugé acceptable. En pratique, on demande que le taux d'erreur soit inférieur au taux

d'erreur le plus élevé des LAP Solystic. Pour cela, nous disposons d'un ensemble d'apprentissage composé d'environ 28000 lettres. C'est sur ce lot d'images que les matrices de confusion ont été calculées, et les coûts ajustés. Un autre lot de 28000 lettres constitue l'ensemble de test sur lequel ont été validés les réglages effectués sur le premier lot. Les proportions de courrier manuscrit et dactylographié, ainsi que des différentes catégories de la hiérarchie postale, sont sensiblement identiques dans l'ensemble d'apprentissage et dans l'ensemble de test.

Dans les graphiques de représentation des performances des LAP, afin de préserver la confidentialité des performances des LAP utilisés, l'origine de l'axe des abscisses représentant le taux de lecture au niveau acheminement (resp. au niveau distribution) aura une valeur de référence noté $\mathcal{T}L_{ref}^A$ (resp. $\mathcal{T}L_{ref}^D$). De même, le taux d'erreur sera exprimé en fonction d'un taux de référence en erreur, noté $\mathcal{T}E_{ref}$.

- le taux de référence en lecture au niveau acheminement $\mathcal{T}L_{ref}^A$ a une valeur supérieure à 80%,
- le taux de référence en lecture au niveau distribution $\mathcal{T}L_{ref}^D$ a une valeur supérieure à 50%,
- le taux de référence en erreur $\mathcal{T}E_{ref}$ a une valeur inférieure à 0.1%.

La figure 3.8 illustre des performances obtenues en phase d'apprentissage au niveau acheminement (figure 3.8(a)), et au niveau distribution (figure 3.8(b)). Sur ces figures, sont représentés différents points de fonctionnement de la combinaison implémentée à partir du modèle défini dans ce chapitre, et nommée *CMCT* pour *Combinaison basée sur le MCT*. Ces points correspondent à des couples (taux de lecture, taux d'erreur) obtenus à partir de différents réglages (R_1, R_2, \dots, R_{15}) des coûts de décision.

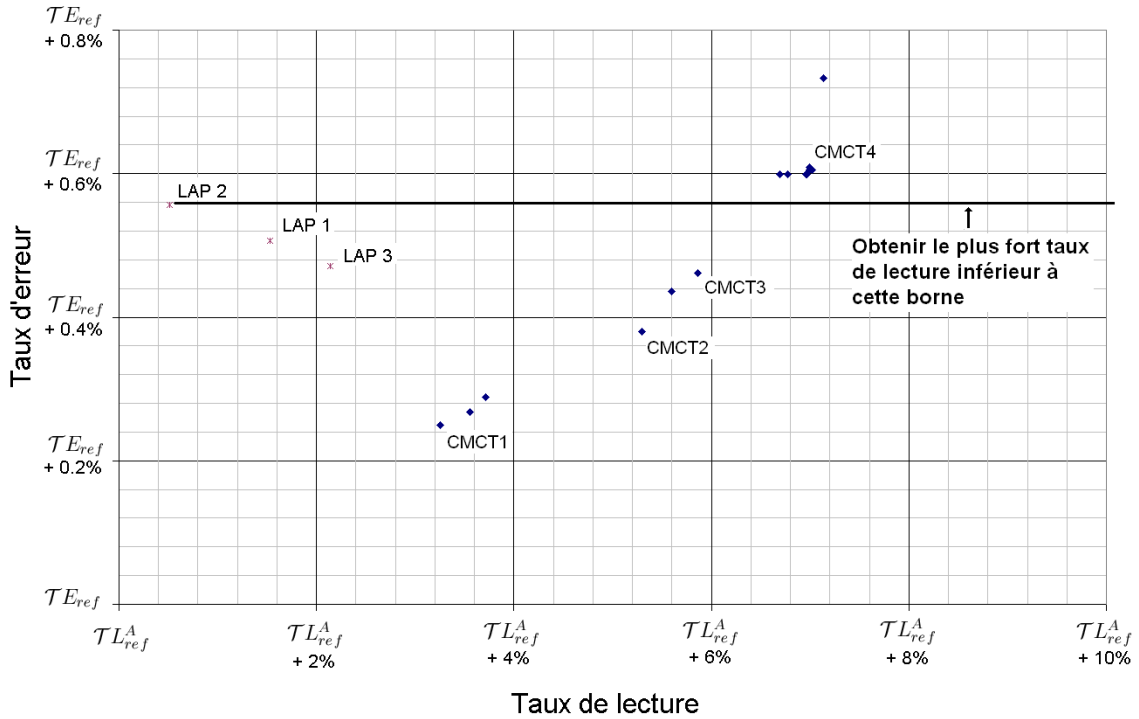
Quatre points de fonctionnement notés $CMCT_i, i \in \{1, 2, 3, 4\}$, sont sélectionnés pour observer le comportement de la combinaison *CMCT* en phase de test, et les comparer à des schémas à base de votes majoritaires définis dans le paragraphe suivant. Ces points sont associés à un réglage particulier : le point $CMCT_1$ (resp. $CMCT_2, CMCT_3, CMCT_4$) est obtenu par le réglage R_1 (resp. R_4, R_6, R_8).

3.4.2 Schémas à base de votes

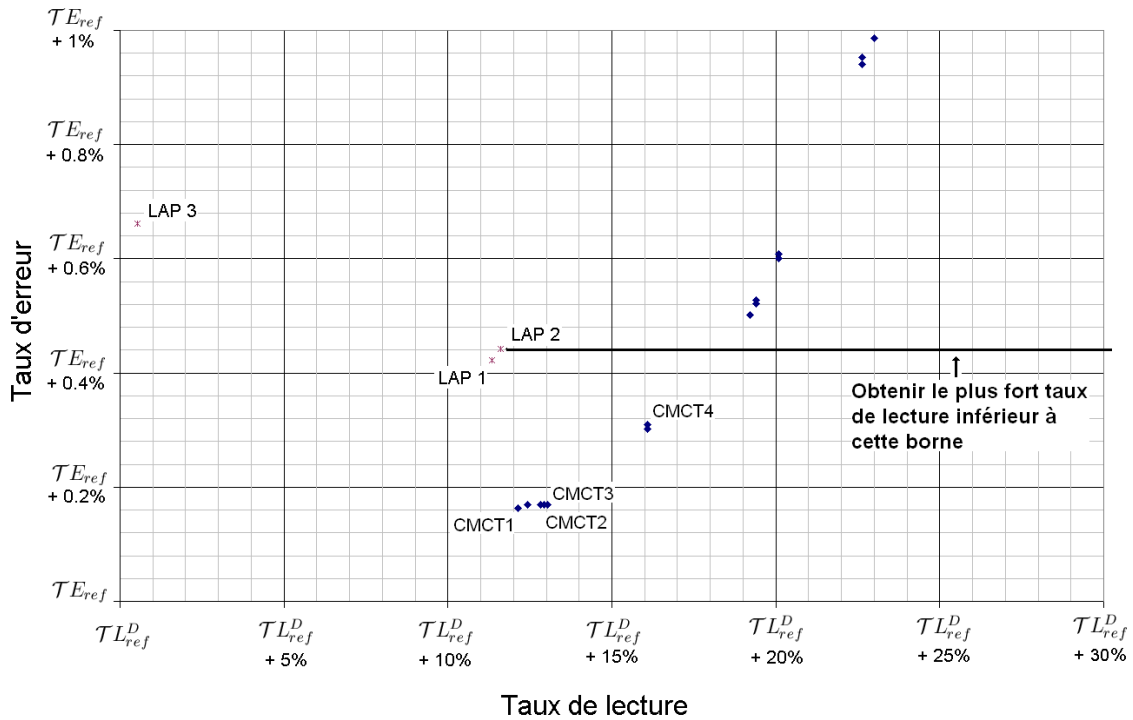
Dans ce paragraphe, des schémas à base de votes sont introduits. Ceux-ci serviront de points de repère afin d'observer les apports de la combinaison crédale décrite dans ce chapitre.

Quand un groupe de classifieurs possède de bonnes performances, le vote à la majorité offre une combinaison facile à implémenter et améliorant toutes les performances individuelles de chaque classifieur [58]. Le processus basé sur ce principe sera noté *Maj*. Afin d'obtenir un taux de lecture plus élevé, une variante consiste à ne choisir la solution majoritaire que si celle-ci est différente d'un rejet complet, et en l'absence de majorité, la solution du LAP jugé le plus performant sera retenu, ici le LAP 1. Cette méthode sera notée *MajLap₁*.

Exemple 3.12. *Si le LAP 1 décide « A_2Rej », le LAP 2 décide « Rej » et le LAP 3 décide « Rej », alors le processus *Maj* décidera « Rej » et le processus *MajLap₁* décidera « A_2Rej ».*



(a) Performances au **niveau acheminement**. Le nombre TL_{ref}^A indique un taux de référence en lecture au niveau acheminement.



(b) Performances au **niveau distribution**. Le nombre TL_{ref}^D indique un taux de référence en lecture au niveau distribution.

Figure 3.8 – Performances des LAP et de différents points de fonctionnement de la combinaison CMCT réalisées sur l'ensemble d'apprentissage.

3.4.3 Phase de test

Les performances au niveau acheminement et distribution des LAP, des schémas à base de votes majoritaires et des points de fonctionnement sélectionnés de la combinaison CMCT sont illustrés sur les figures 3.9(a) et 3.9(b).

Un net gain est obtenu par rapport aux performances individuelles des LAP au niveau acheminement. Ce gain est plus réduit au niveau distribution, notamment en terme de gain de lectures, c'est-à-dire en terme de gain de bonnes reconnaissances. Rappelons que lorsque nous parlons de courrier lu, nous parlons de courrier bien reconnu.

Dans les deux cas, la combinaison CMCT démontre une certaine robustesse dans son comportement, puisque les points de fonctionnement restent dans la même « zone de performance » sur les deux lots de courrier, et cela, malgré des LAP aux performances en distribution aux variations sensibles. Néanmoins, la sélection d'un point de fonctionnement pour la combinaison n'est pas aisée. En effet, sur le lot de courrier d'apprentissage (figure 3.8), les quatre points $CMCT_i$ ont un taux d'erreur acceptable en distribution, toutefois le meilleur point en distribution, $CMCT_4$, possède un taux d'erreur au niveau acheminement trop élevé. Ainsi, le point sélectionné sera le point $CMCT_3$, qui est acceptable aux niveaux acheminement et distribution. Cependant, comme nous pouvons l'observer sur l'ensemble de test (figure 3.9), ce point peut parfois dépasser légèrement la limite du taux d'erreur acceptable au niveau acheminement. Par conséquent, il serait plus prudent de choisir un réglage un peu plus sévère sur le taux d'erreur, et de choisir le point $CMCT_2$. Il aurait été préférable de chercher un point intermédiaire entre $CMCT_2$ et $CMCT_3$, afin de conserver des gains de lecture en acheminement, mais comme la remarque 3.3 le souligne, ceci n'est actuellement pas possible.

3.4.4 Remarques

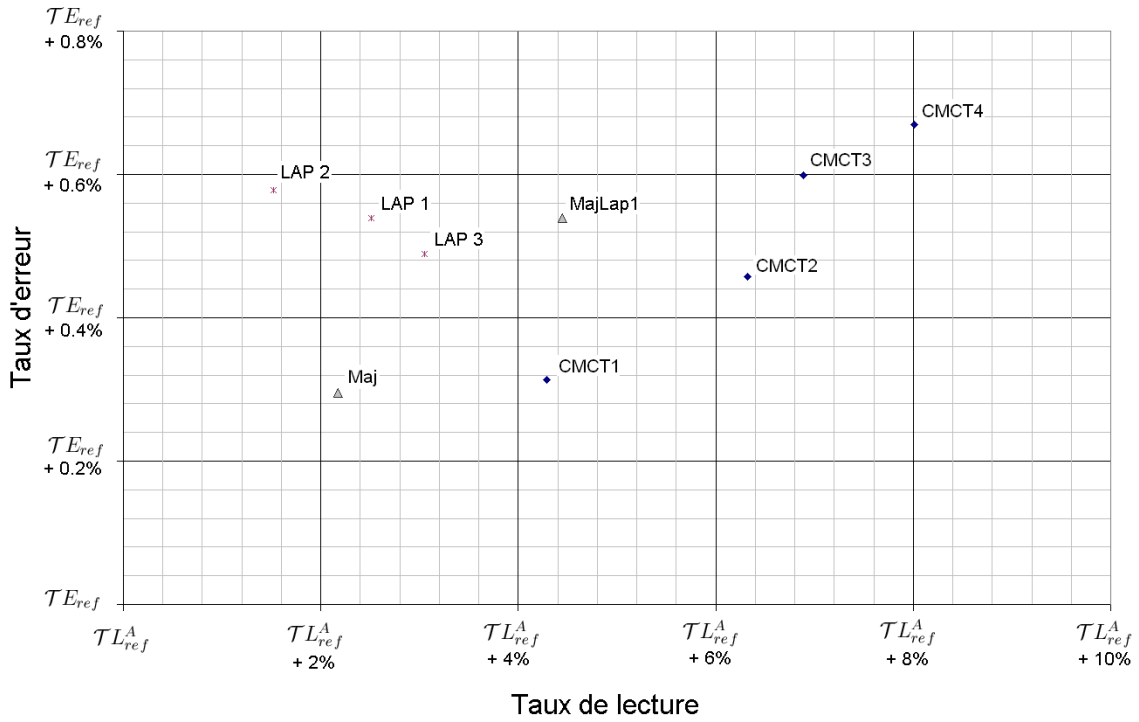
Remarque 3.3 (Des performances « par sauts »). *Nous pouvons remarquer certains « sauts » entre les différents points de fonctionnement de la combinaison CMCT (figure 3.8), plus précisément illustrés sur les figures 3.10 et 3.11.*

Ces sauts proviennent de l'affectation qui caractérise, pour chaque LAP, toute décision d'une même catégorie de manière identique. Nous appellerons situation, chacun de ces triplets similaires de décisions.

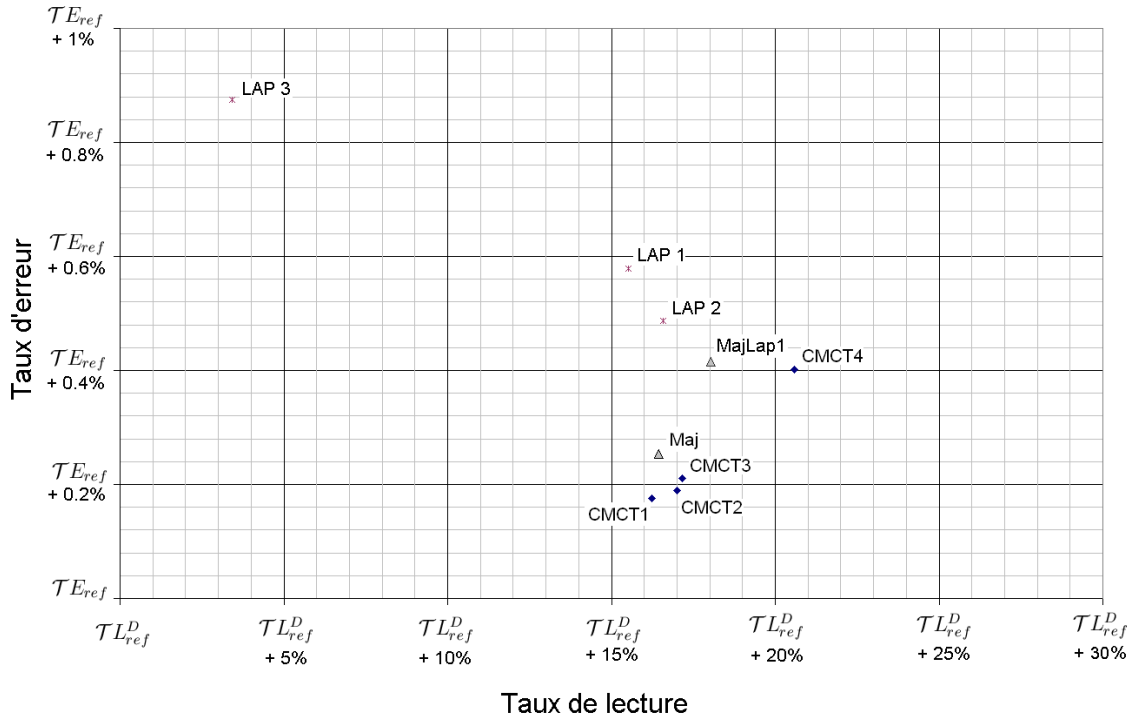
Par exemple, en distribution dactylographiée (figure 3.11(a)), le saut entre les réglages R_{13} et R_{14} , ne gagnant que peu de bonnes reconnaissances par rapport aux nombres d'erreurs apportées, est dû principalement à l'acceptation de la décision distribution proposée par le LAP 3, lorsque les deux LAP 1 et 2 ont fourni une même décision de niveau acheminement en accord avec la décision du LAP 3, autrement dit, à l'acceptation de la décision distribution proposée par le LAP 3 dans la situation suivante :

- le LAP 1 décide : « $A_i Rej$ »,
- le LAP 2 décide : « $A_i Rej$ »,
- le LAP 3 décide : « $A_i R_k$ »,

où A_i est un acheminement, et R_k une voie dans A_i . Avec le réglage R_{13} , la décision



(a) Performances au **niveau acheminement**.



(b) Performances au **niveau distribution**.

Figure 3.9 – Performances des LAP, des schémas à base de votes majoritaires et des points de fonctionnement de la combinaison CMCT sélectionnés, réalisées sur l'ensemble de test.

fournie par la combinaison est « $A_i\text{Rej}$ ». Avec le réglage R_{14} , la décision fournie est « A_iR_k ». Entre le réglage R_{13} et le réglage R_{14} , le coût de rejet distribution a été augmenté.

En distribution manuscrite (figure 3.11(b)), l'acceptation de la décision du LAP 3 dans cette même situation est à l'origine du deuxième saut entre les réglages R_8 et R_9 . Le premier saut, entre les réglages R_6 et R_7 , est principalement dû à l'acceptation de la décision distribution proposée par les LAP 1 et 2 dans les situations suivantes :

- le LAP 1 décide : « A_iR_k »,
- le LAP 2 décide : « $A_i\text{Rej}$ »,
- le LAP 3 décide : « $A_i\text{Rej}$ »,

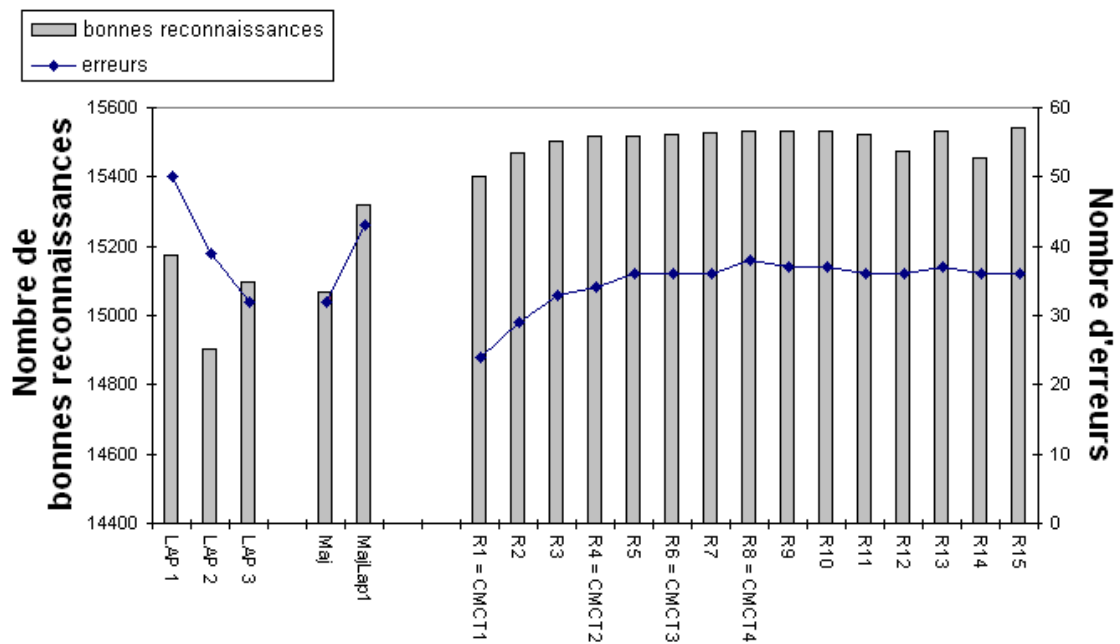
ou,

- le LAP 1 décide : « $A_i\text{Rej}$ »,
- le LAP 2 décide : « A_iR_k »,
- le LAP 3 décide : « $A_i\text{Rej}$ ».

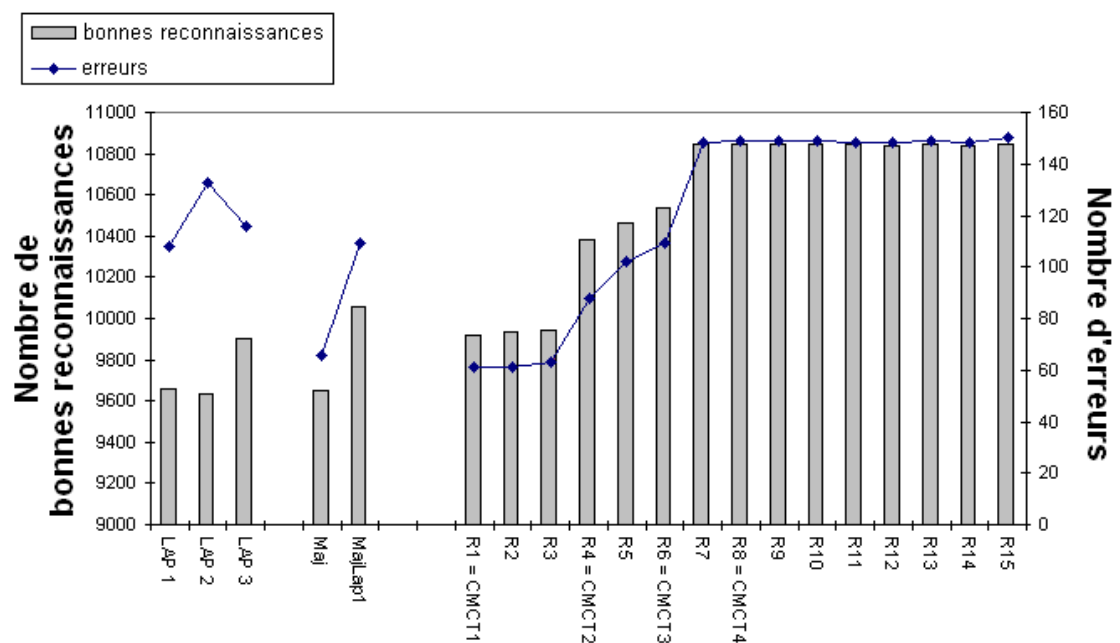
Remarquons que ces observations indiquent que la difficulté actuelle de la combinaison en distribution n'est pas due à un conflit direct de choix entre deux rues proposées, mais à l'acceptation d'une distribution proposée alors qu'aucune autre solution n'est proposée. Dans l'affectation postale employée, une information fournie par les matrices de confusion quant à l'invalidité d'une adresse lors d'un rejet, a déjà été prise en compte. Néanmoins, pour discriminer des lettres à l'intérieur des lots de courrier conduisant à ces situations, il faudra tenir compte de nouvelles informations. Nous donnons des éléments dans la conclusion de ce chapitre.

Remarque 3.4 (Problèmes du LAP 3 en distribution). On observe que le LAP 3 est moins performant en distribution que les LAP 1 et 2. En réalité, outre le fait qu'il reconnaît moins d'intitulés géographiques et avec moins de réussite que les LAP 1 et 2, ce dernier ne reconnaît pas les distributions spéciales, c'est-à-dire les boîtes postales. Dans le meilleur des cas, il effectue un rejet en distribution lorsqu'il est confronté à des envois d'une telle nature. Cette information n'est pas explicitement prise en compte dans le présent modèle.

Remarque 3.5 (Concernant le conflit). À partir du postulat selon lequel une prise de décision à partir d'une fonction de croyance est d'autant plus sûre que celle-ci alloue une masse faible au vide, une variante de ce modèle de combinaison a consisté à poser des seuils sur le conflit, à partir desquels on acceptait de prendre une décision autre que le rejet. Autrement dit, si la masse sur le conflit était supérieure à un seuil, un rejet total était décidé. En faisant varier ces seuils, nous n'avons pas amélioré les points de fonctionnement obtenus avec le modèle présent avec différents réglages des coûts.

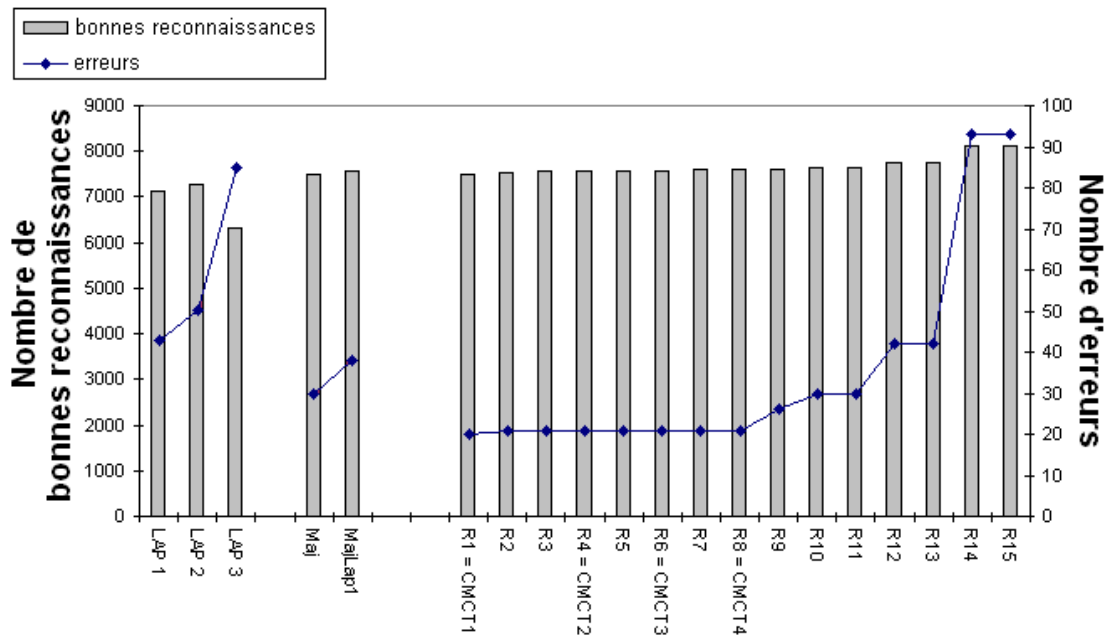


(a) En acheminement dactylographié.

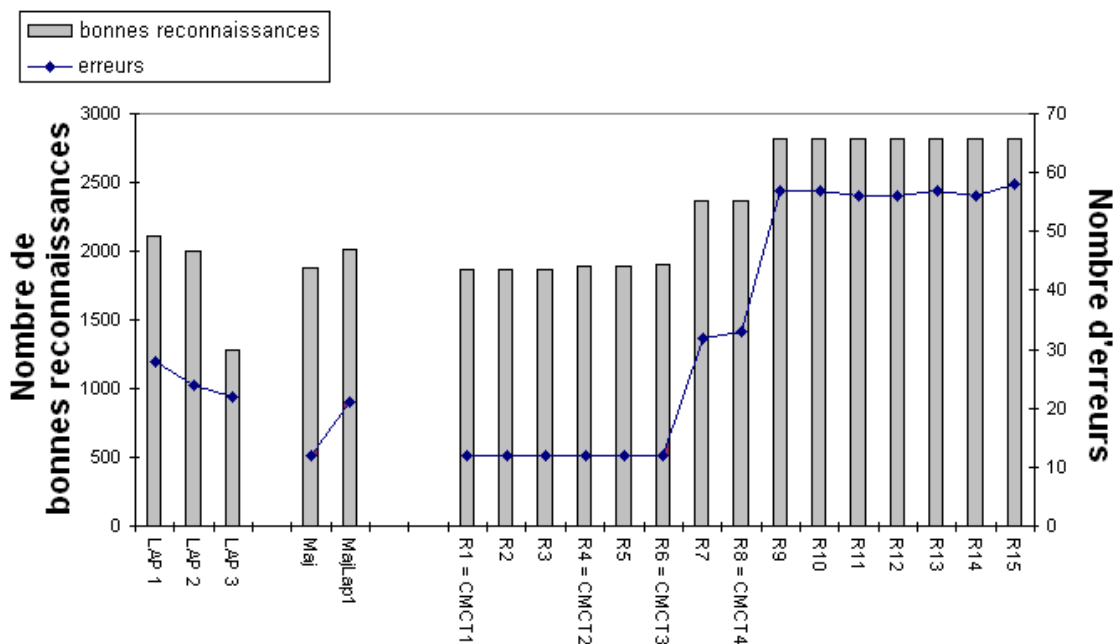


(b) En acheminement manuscrit.

Figure 3.10 – Nombres de bonnes reconnaissances et d’erreurs pour chaque LAP et chaque réglage de la combinaison sur le lot de courrier d’apprentissage.



(a) En distribution dactylographiée.



(b) En distribution manuscrite.

Figure 3.11 – Nombres de bonnes reconnaissances et d'erreurs pour chaque LAP et chaque réglage de la combinaison sur le lot de courrier d'apprentissage.

3.5 Conclusion

Dans ce chapitre, un modèle de fusion de décisions postales basées sur la théorie des fonctions de croyance a été présenté. Ses principales caractéristiques sont les suivantes :

- il s'appuie sur une affectation des masses basée sur une hiérarchie et une catégorisation des décisions postales dont une des spécificités est de pouvoir s'exprimer à plusieurs niveaux. Remarquons à nouveau que cette affectation peut aussi s'employer dans le cas de décisions s'exprimant uniquement sur les singletons de l'univers. Dans ce cas, la hiérarchie s'interprète comme un lien entre les décisions qui justifie une affectation des masses intermédiaire entre une affectation Bayésienne et une affectation de Xu et al ;
- la méthode de prise de décision adoptée permet un ajustement de la combinaison aux contraintes de taux de lecture et taux d'erreur attendues par l'utilisateur ;
- ce modèle de combinaison peut s'appliquer à tout type de LAP d'origine Solystic ou extérieure. Ainsi cette combinaison, peut donner naissance à un nouveau produit, commercialisable indépendamment des LAP Solystic ;
- bien que n'employant que peu d'outils dans la manipulation des fonctions de croyance, ce modèle fournit d'ores et déjà des performances intéressantes.

Pour améliorer cette combinaison, au moins deux métaconnaissances sur les LAP peuvent être prises en compte :

1. des informations de fiabilité des LAP, pouvant dépendre de contextes ;
2. des informations de dépendances entre les LAP Solystic 1 et 2.

D'un autre côté, les LAP 1 et 2 pourront fournir en plus de leurs décisions, des scores de confiance calculés à partir des algorithmes employés. Ces informations supplémentaires, intrinsèques à chaque image d'envoi, devraient améliorer la fusion proposée.

Dans le prochain chapitre, de nouveaux mécanismes de correction de fonctions de croyance sont introduits à partir d'hypothèses réalisées sur la fiabilité contextuelle d'une source d'information, ou sur les états mêmes dans lesquels peut se trouver une source d'information.

Mécanismes de correction de fonction de croyance

4.1 Introduction

Dans le chapitre précédent, une méthode d'affectation a été introduite afin de convertir les décisions des LAP en une fonction de croyance, à partir de leurs matrices de confusion. Par cette origine, ces fonctions de croyance reflètent des connaissances statistiques toujours disponibles. Dans ce chapitre, nous nous intéressons au cas où des informations complémentaires relatives à la fiabilité des LAP sont disponibles.

Dans le MCT, la prise en compte de la fiabilité d'une source est classiquement réalisée par l'opération d'affaiblissement rappelée au paragraphe 2.4.8. Comme Smets l'a montré [98], cette opération n'est pas *ad hoc*, mais peut être obtenue à partir d'un simple modèle de la fiabilité. Dans ce modèle, la source peut être dans deux états de fiabilité :

- soit elle est fiable, et dans ce cas l'information qu'elle fournit est acceptée ;
- soit elle est non fiable, et l'information qu'elle fournit est rejetée.

En pratique, cette fiabilité n'est pas connue avec certitude, d'où l'existence de la constante $\alpha \in [0, 1]$, telle que $(1 - \alpha)$ représente le degré de fiabilité de la source. Chaque information disponible peut alors être modélisée par une fonction de croyance, définie soit sur le cadre de discernement Ω de la variable d'intérêt, soit sur le cadre de discernement \mathcal{R} composé des états de fiabilité de la source.

Dans cette opération d'affaiblissement, la connaissance sur la fiabilité d'une source est modélisée par un unique nombre. Dans certains cas, cependant, des connaissances plus fines peuvent être disponibles. En particulier, la fiabilité d'une source d'information peut dépendre de la valeur prise par la variable d'intérêt. Par exemple, dans notre application postale, nous avons vu que le LAP 3 ne sait pas reconnaître les boîtes postales ; en revanche il possède une fiabilité non nulle dans la détection des intitulés géographiques. De même, en diagnostic médical, de part son expérience passée, son entraînement ou sa spécialité, un médecin peut être particulièrement compétent pour diagnostiquer certaines pathologies, et beaucoup moins pour d'autres.

Afin de prendre en compte de telles informations, nous proposons une extension de cet affaiblissement dans laquelle l'utilisateur peut exprimer une confiance, conditionnellement à la valeur ou à un ensemble de valeurs pris par la variable d'intérêt. La combinaison de ces informations conduit au développement d'un nouvel outil, appelé *affaiblissement contextuel* [70, 71], contrôlé non plus par un seul taux

d'affaiblissement, mais par un vecteur de taux d'affaiblissement décrivant la fiabilité de la source dans chacun des contextes. Nous montrons de plus qu'en généralisant une méthode d'apprentissage automatique des taux d'affaiblissement introduite par Elouedi et al. [40], ce vecteur peut aussi être appris à partir d'un ensemble de données étiquetées [69, 71].

D'autre part, nous pouvons remarquer que l'opération d'affaiblissement permet seulement d'atténuer une fonction de croyance, alors qu'au contraire, il pourrait être souhaité de renforcer une source d'information lorsque celle-ci est jugée trop prudente. Ainsi, dans [68], nous avons défini un mécanisme général de correction permettant d'affaiblir ou de renforcer une fonction de croyance. Dans le paragraphe 4.7 de ce chapitre, nous définissons un cadre plus général permettant de développer des mécanismes de correction de fonctions de croyance plus précis. Par exemple, au lieu de vouloir seulement affaiblir ou renforcer une source, il peut aussi être souhaité de la contredire. De même, nous montrons qu'une source peut être renforcée par différents mécanismes plus ou moins engagés.

Ainsi, nous montrons dans ce chapitre que l'opération d'affaiblissement, que l'on nommera *affaiblissement classique*, est en réalité un mécanisme de correction de fonction de croyance particulier : en fonction d'hypothèses posées sur la fiabilité de la source et les états de la source, différents mécanismes de correction de l'information fournie par cette source peuvent être définis. Le schéma commun à tous ces mécanismes est illustré sur la figure 4.1.

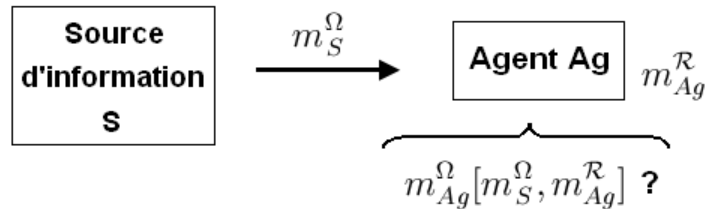


Figure 4.1 – Représentation des croyances entrant en jeu dans la correction de l'information fournie par la source.

Les preuves les plus longues de ce chapitre sont regroupées dans l'annexe A de ce mémoire.

4.2 Extension vide et déconditionnement dans le cas d'un espace produit

Les concepts d'extension vide et de déconditionnement dans le cas d'un espace produit jouent un rôle important dans la définition de l'opération d'affaiblissement ainsi que dans les autres mécanismes de correction que nous avons développés. Ces concepts sont illustrés dans les paragraphes suivants.

4.2.1 Marginalisation et extension vide dans le cas d'un espace produit

Une fonction de masse définie sur un espace produit $\Omega \times \Theta$ peut être marginalisée sur Ω , par le transfert de chaque masse $m^{\Omega \times \Theta}(B)$, $B \subseteq \Omega \times \Theta$, vers sa projection sur Ω :

$$m^{\Omega \times \Theta \downarrow \Omega}(A) = \sum_{\{B \subseteq \Omega \times \Theta \mid \text{Proj}(B \downarrow \Omega) = A\}} m^{\Omega \times \Theta}(B), \forall A \subseteq \Omega \quad (4.1)$$

où $\text{Proj}(B \downarrow \Omega)$ dénote la projection de B sur Ω .

Il n'est habituellement pas possible de retrouver la fonction de masse originale $m^{\Omega \times \Theta}$ à partir de sa marginalisation $m^{\Omega \times \Theta \downarrow \Omega}$ sur Ω . Cependant, la fonction de masse la moins informative telle que sa projection sur Ω donne $m^{\Omega \times \Theta \downarrow \Omega}$, peut être calculée. Cette opération est appelée *extension vide* de m^Ω sur l'espace produit $\Omega \times \Theta$ [98], notée $m^{\Omega \uparrow \Omega \times \Theta}$, et, définie par :

$$m^{\Omega \uparrow \Omega \times \Theta}(B) = \begin{cases} m^\Omega(A) & \text{si } B = A \times \Theta, A \subseteq \Omega \\ 0 & \text{sinon.} \end{cases} \quad (4.2)$$

La figure 4.2 illustre, d'une part, le sous-ensemble A de Ω sur lequel est transférée, après marginalisation sur Ω , une masse initialement allouée à un sous-ensemble B de $\Omega \times \Theta$, et, d'autre part, le sous-ensemble B de $\Omega \times \Theta$ sur lequel est transférée, après extension vide sur $\Omega \times \Theta$, une masse initialement allouée à un sous-ensemble A de Ω .

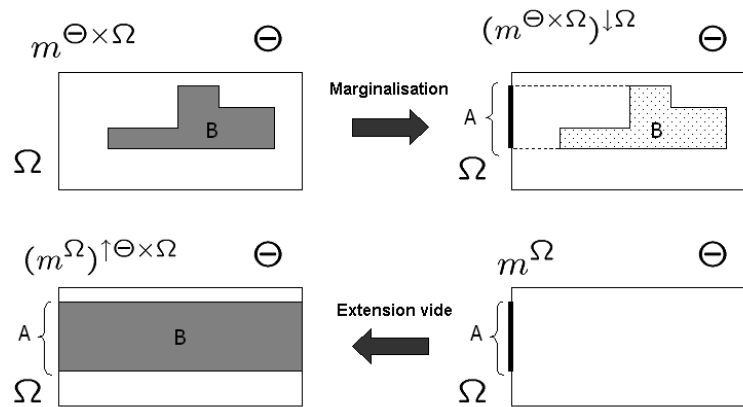


Figure 4.2 – Illustration [27] d'un transfert de masse par marginalisation (en haut), et par extension vide (en bas), dans le cas d'un espace produit.

4.2.2 Conditionnement et déconditionnement dans le cas d'un espace produit

Si $m^{\Omega \times \Theta}$ est définie sur un espace produit $\Omega \times \Theta$, D étant un sous-ensemble de Θ , la fonction de masse conditionnelle $m^\Omega[D]$ est définie par la combinaison de $m^{\Omega \times \Theta}$

avec $m_D^{\Theta \uparrow \Omega \times \Theta}$ marginalisée sur Ω :

$$m^\Omega[D] = \left(m^{\Omega \times \Theta} \ominus m_D^{\Theta \uparrow \Omega \times \Theta} \right) \downarrow^\Omega. \quad (4.3)$$

Réciproquement, supposons $m^\Omega[D]$ connue, i.e., la croyance d'un agent sur Ω dans un contexte où la valeur sur Θ est dans D . Généralement, il y a plusieurs fonctions de masse définies sur $\Omega \times \Theta$, dont le conditionnement sur D est $m^\Omega[D]$. Parmi ces fonctions, la moins informative est définie par :

$$m^\Omega[D] \uparrow^{\Omega \times \Theta} (A \times D \cup \Omega \times \bar{D}) = m^\Omega[D](A), \quad \forall A \subseteq \Omega. \quad (4.4)$$

$m^\Omega[D] \uparrow^{\Omega \times \Theta}$ constitue le déconditionnement de $m^\Omega[D]$ sur l'espace produit $\Omega \times \Theta$.

La figure 4.3 illustre, d'une part le sous-ensemble A de Ω sur lequel est transférée, après conditionnement sur Ω , une masse initialement allouée à un sous-ensemble B de $\Omega \times \Theta$, et, d'autre part, le sous-ensemble B de $\Omega \times \Theta$ sur lequel est transférée, après déconditionnement sur $\Omega \times \Theta$, une masse initialement allouée à un sous-ensemble A de Ω . Ce sous-ensemble B de $\Omega \times \Theta$ est le plus grand des sous-ensembles de $\Omega \times \Theta$, qui voient leur masse transférée à A après conditionnement sur Ω .

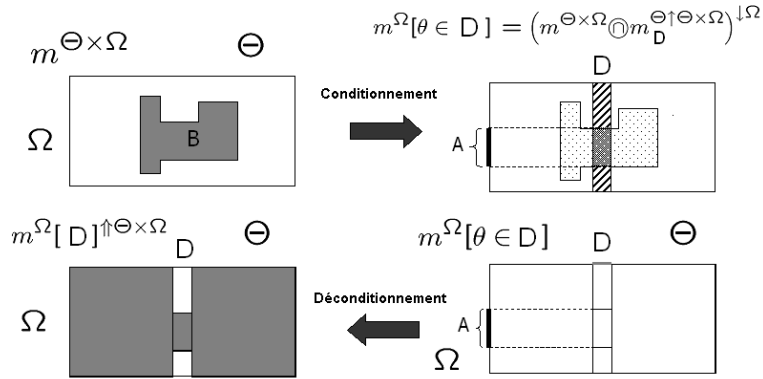


Figure 4.3 – Illustration [27] d'un transfert de masse par conditionnement (en haut), et par déconditionnement (en bas), dans le cas d'un espace produit.

4.3 Affaiblissement classique

Dans ce modèle, l'information sur la *fiabilité* de S est quantifiée par une fonction de masse $m_{Ag}^{\mathcal{R}}$ sur $\mathcal{R} = \{R, NR\}$, où R signifie « la source est fiable », et NR signifie « la source est non fiable » [98].

Le modèle de la fiabilité est définie de la manière suivante :

- si S est fiable, l'agent Ag accepte l'information fournie par S :

$$m_{Ag}^\Omega[R] = m_S^\Omega, \quad (4.5)$$

où, $m_{Ag}^\Omega[\{R\}]$ est plus simplement notée $m_{Ag}^\Omega[R]$, ce qui sera le cas dans tout ce mémoire ;

- si S est non fiable, l'agent Ag ne peut pas prendre en compte l'information fournie par S , et sa connaissance demeure l'ignorance totale :

$$m_{Ag}^{\Omega}[NR](\Omega) = 1 . \quad (4.6)$$

En supposant, $m_{Ag}^{\mathcal{R}}$ de la forme suivante :

$$\begin{cases} m_{Ag}^{\mathcal{R}}(\{R\}) &= 1 - \alpha \\ m_{Ag}^{\mathcal{R}}(\mathcal{R}) &= \alpha, \end{cases} \quad (4.7)$$

avec $\alpha \in [0, 1]$, deux fonctions de masse non vides sont à combiner : $m_{Ag}^{\mathcal{R}}$ et $m_{Ag}^{\Omega}[R]$. Ces deux fonctions de masse ne s'expriment pas sur le même référentiel. Il faut donc les ramener sur un référentiel commun : l'espace produit.

En étendant $m_{Ag}^{\mathcal{R}}$ à $\Omega \times \mathcal{R}$, et en déconditionnant $m_{Ag}^{\Omega}[R]$ sur le même espace produit, ces fonctions peuvent être conjonctivement combinées. Celles-ci peuvent en effet être considérées comme distinctes, puisque le modèle de la fiabilité étant donné, l'agent Ag est libre de choisir n'importe quel taux d'affaiblissement α . Enfin, le résultat de cette combinaison est marginalisé sur Ω , afin d'obtenir l'information sur Ω . Au total, les informations disponibles sont fusionnées de la manière suivante :

$$m_{Ag}^{\Omega}[m_S^{\Omega}, m_{Ag}^{\mathcal{R}}] = \left(m_{Ag}^{\Omega}[R]^{\uparrow\Omega \times \mathcal{R}} \odot m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}} \right)^{\downarrow\Omega} , \quad (4.8)$$

La fonction résultante $m_{Ag}^{\Omega}[m_S^{\Omega}, m_{Ag}^{\mathcal{R}}]$, où les crochets indiquent le corpus des évidences, c'est-à-dire ce qui est connu de l'agent Ag , dépend seulement de m_S^{Ω} et α , comme le détaille la remarque 4.1. Ainsi, la fonction de masse affaiblie $m_{Ag}^{\Omega}[m_S^{\Omega}, m_{Ag}^{\mathcal{R}}]$ peut se noter ${}^{\alpha}m_S^{\Omega}$, ou simplement ${}^{\alpha}m$. Son expression est la suivante :

$$\begin{cases} {}^{\alpha}m(A) &= (1 - \alpha)m(A), \quad \forall A \subset \Omega, \\ {}^{\alpha}m(\Omega) &= (1 - \alpha)m(\Omega) + \alpha, \end{cases} \quad (4.9)$$

ou, plus simplement :

$${}^{\alpha}m = (1 - \alpha)m + \alpha m_{\Omega} . \quad (4.10)$$

Cette opération, appelée *affaiblissement* (en anglais, *discounting*), a été introduite de manière intuitive par Shafer dans son livre [88, page 251]. La justification formelle présentée ici est due à Smets [98].

Remarque 4.1. *Étant donné le caractère instructif de cette preuve pour la suite, les détails des calculs conduisant à l'équation (4.9) sont repris dans cette remarque. Pour tout $A \subseteq \Omega$, comme illustré sur la figure 4.4 :*

$$m_{Ag}^{\Omega}[R]^{\uparrow\Omega \times \mathcal{R}}(A \times \{R\} \cup \Omega \times \{NR\}) = m_{Ag}^{\Omega}[R](A) = m_S^{\Omega}(A). \quad (4.11)$$

L'extension vide de $m_{Ag}^{\mathcal{R}}$ se calcule facilement :

$$\begin{cases} m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}}(\Omega \times \{R\}) &= 1 - \alpha \\ m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}}(\Omega \times \mathcal{R}) &= \alpha. \end{cases} \quad (4.12)$$

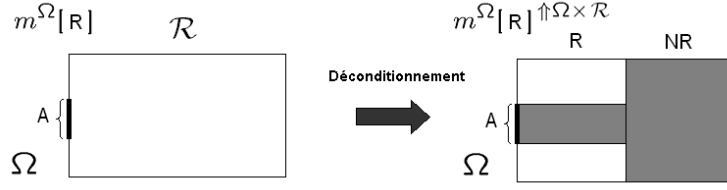


Figure 4.4 – Illustration du transfert de masse par déconditionnement d'une masse définie sur Ω conditionnelle à R sur l'espace produit $\Omega \times \mathcal{R}$.

Pour tout $A \subseteq \Omega$:

$$\begin{cases} (A \times \{R\} \cup \Omega \times \{NR\}) \cap \Omega \times \{R\} &= A \times \{R\} \\ (A \times \{R\} \cup \Omega \times \{NR\}) \cap \Omega \times \mathcal{R} &= A \times \{R\} \cup \Omega \times \{NR\}, \end{cases} \quad (4.13)$$

et donc :

$$\begin{cases} m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}} \ominus m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}}(A \times \{R\}) &= (1 - \alpha) m_S^\Omega(A) \\ m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}} \ominus m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}}(A \times \{R\} \cup \Omega \times \{NR\}) &= \alpha m_S^\Omega(A). \end{cases} \quad (4.14)$$

Après projection sur Ω , on obtient :

$$\begin{cases} \alpha m_{Ag}^\Omega(A) &= (1 - \alpha) m_S^\Omega(A), \quad \forall A \subset \Omega \\ \alpha m_{Ag}^\Omega(\Omega) &= (1 - \alpha) m_S^\Omega(\Omega) + \sum_{A \subseteq \Omega} \alpha m_S^\Omega(A) \\ &= (1 - \alpha) m_S^\Omega(\Omega) + \alpha. \end{cases} \quad (4.15)$$

Remarque 4.2. Un affaiblissement peut aussi s'exprimer par la fonction d'impli-cabilité :

$$\alpha b(A) = \begin{cases} (1 - \alpha)b(A) & \text{si } A \neq \Omega \\ 1 & \text{sinon.} \end{cases} \quad (4.16)$$

Remarque 4.3. αm est une généralisation de m :

$$\alpha m(A) = \sum_{B \subseteq \Omega} \alpha G(A, B) m(B), \quad (4.17)$$

où αG est une matrice de généralisation définie par :

$$\alpha G(A, B) = \begin{cases} 1 - \alpha & \text{si } A = B \neq \Omega, \\ \alpha & \text{si } A = \Omega \text{ et } B \subset A, \\ 1 & \text{si } A = B = \Omega \\ 0 & \text{sinon.} \end{cases} \quad (4.18)$$

La matrice αG a la forme suivante :

$$\alpha G = \begin{pmatrix} 1 - \alpha & 0 & \dots & 0 & 0 \\ 0 & 1 - \alpha & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & 1 - \alpha & 0 \\ \alpha & \alpha & \dots & \alpha & 1 \end{pmatrix} \quad (4.19)$$

Rappelons que $\alpha G(A, B)$ représente la fraction de la masse $m(B)$ qui est transférée à A lors de l'application de l'affaiblissement.

Remarque 4.4. La fonction de masse ${}^\alpha m$ est le résultat de la combinaison disjunctive de m avec m_0 définie par :

$$\begin{cases} m_0(\emptyset) &= 1 - \alpha \\ m_0(\Omega) &= \alpha . \end{cases} \quad (4.20)$$

En effet, $b_0(A) = 1 - \alpha$, $\forall A \subset \Omega$, et $b_0(\Omega) = 1$, d'où $b_0 b = {}^\alpha b$.

Remarque 4.5. Si $m_{Ag}^{\mathcal{R}}$ est une fonction de masse Bayésienne :

$$\begin{cases} m_{Ag}^{\mathcal{R}}(\{R\}) &= 1 - \alpha , \\ m_{Ag}^{\mathcal{R}}(\{NR\}) &= \alpha , \end{cases} \quad (4.21)$$

le résultat de l'affaiblissement demeure identique [98].

Exemple 4.1. Comme dans [40], considérons un problème simplifié de reconnaissance de cible, dans lequel les objets appartiennent uniquement aux trois classes suivantes : avion ($\omega_1 \equiv a$), hélicoptère ($\omega_2 \equiv h$) et missile ($\omega_3 \equiv r$). Soit $\Omega = \{a, h, r\}$. Supposons qu'un capteur ait fourni la fonction de masse suivante au regard d'une cible : $m(\{a\}) = 0.5$, $m(\{r\}) = 0.5$, signifiant une hésitation entre classer cette cible comme un avion ou comme un missile. Supposons qu'un agent reçoive cette information et ait un degré de croyance égale à $1 - \alpha$ dans le fait que ce capteur soit fiable. Un taux d'affaiblissement α est alors appliquée à la fonction de masse m fournie par le capteur. Avec $\alpha = 0.4$, la fonction de masse de l'agent devient : ${}^\alpha m(\{a\}) = 0.5(1 - \alpha) = 0.3$, ${}^\alpha m(\{r\}) = 0.5(1 - \alpha) = 0.3$, ${}^\alpha m(\Omega) = \alpha = 0.4$.

4.4 Affaiblissement contextuel

Dans ce paragraphe, l'opération d'affaiblissement est étendue afin de permettre la prise en compte d'informations plus fines au regard de la fiabilité de la source dans différents contextes.

4.4.1 Expression de la fiabilité

Dans ce paragraphe, des informations sur la fiabilité de la source S , conditionnelles à chaque $\omega_k \in \Omega$, c'est-à-dire relatives à chaque contexte dans lequel la vérité est ω_k , sont supposées connues par Ag . Ainsi, la fiabilité sur S s'exprime ici, non plus par une fonction de masse non conditionnelle comme dans l'équation 4.7, mais par K fonctions de masse conditionnelles $m_{Ag}^{\mathcal{R}}[\omega_k]$, $k = 1, \dots, K$, de la forme suivante :

$$\begin{cases} m_{Ag}^{\mathcal{R}}[\omega_k](\{R\}) &= \beta_k, \\ m_{Ag}^{\mathcal{R}}[\omega_k](\mathcal{R}) &= \alpha_k, \end{cases} \quad (4.22)$$

où $\beta_k = 1 - \alpha_k$ représente le degré de fiabilité de la source sachant que la vérité est ω_k , c'est-à-dire sachant que la variable d'intérêt x a pris la valeur ω_k . Rappelons que, comme précédemment, $m_{Ag}^{\mathcal{R}}[\{\omega_k\}]$ est notée $m_{Ag}^{\mathcal{R}}[\omega_k]$ afin de faciliter la lecture.

Le modèle de la fiabilité est conservé. Comme précédemment, l'information fournie par la source est conservée, si celle-ci est fiable, totalement écartée sinon.

Ainsi, $K + 1$ fonctions de masse non vides sont disponibles :

- les K fonctions de masse conditionnelles $m_{Ag}^{\mathcal{R}}[\omega_k]$ exprimant la fiabilité de la source, et
- la fonction de masse $m_{Ag}^{\Omega}[R] = m_S^{\Omega}$ de l'agent Ag au regard de la vérité de x , qui est conditionnelle au fait que la source soit fiable.

Toutes ces fonctions peuvent être déconditionnées sur $\Omega \times \mathcal{R}$, et, supposées distinctes : le fait d'être fiable ou non sur la détection d'un élément de Ω n'a pas d'influence sur la fiabilité de la source dans les autres contextes. Les déconditionnements peuvent alors être combinés conjonctivement, et le résultat peut être marginalisé sur Ω . Cette opération peut être formellement notée :

$$m_{Ag}^{\Omega} [m_S^{\Omega}, m_{Ag}^{\mathcal{R}}[\omega_1], \dots, m_{Ag}^{\mathcal{R}}[\omega_K]] \\ = (m_{Ag}^{\Omega}[R]^{\uparrow\Omega \times \mathcal{R}} \odot m_{Ag}^{\mathcal{R}}[\omega_1]^{\uparrow\Omega \times \mathcal{R}} \odot \dots \odot m_{Ag}^{\mathcal{R}}[\omega_K]^{\uparrow\Omega \times \mathcal{R}})^{\downarrow\Omega}. \quad (4.23)$$

Le résultat de cette combinaison ne dépend que de m_S^{Ω} et du vecteur $\alpha = (\alpha_1, \dots, \alpha_K)$ composé des taux d'affaiblissement. Ainsi, il peut être noté αm_S^{Ω} , ou plus simplement αm , ce qui demeure consistant avec la notation employée pour désigner l'affaiblissement classique (équation 4.9). La transformation de m en αm est appelée *affaiblissement contextuel* de vecteur d'affaiblissement α .

4.4.2 Expression de αm

Proposition 4.1. *Soient $m \in \mathcal{M}^{\Omega}$ et $\alpha = (\alpha_1, \dots, \alpha_K) \in [0, 1]^K$. L'expression de l'affaiblissement contextuel de m par le vecteur d'affaiblissement α , est la suivante :*

$$\alpha m(A) = \sum_{B \subseteq \Omega} \alpha G(A, B) m(B), \quad \forall A \subseteq \Omega, \quad (4.24)$$

avec :

$$\alpha G(A, B) = \begin{cases} \prod_{\omega_k \in A \setminus B} \alpha_k \prod_{\omega_\ell \in \bar{A}} \beta_\ell & \text{si } B \subseteq A, \\ 0 & \text{sinon,} \end{cases} \quad (4.25)$$

où, suivant la convention mathématique classique, un produit de termes portant sur des indices vides est égal à 1.

Preuve : voir paragraphe A.1. □

Les coefficients $\alpha G(A, B)$ pour tout $A, B \subseteq \Omega$ définissent une matrice de généralisation : $\alpha G(A, B)$ représente la fraction de la masse $m(B)$ transférée à A . D'après l'équation 4.25, cette fraction augmente avec :

- la plausibilité α_k que la source soit non fiable, sachant que la vérité est ω_k , pour tout $\omega_k \in A \setminus B$;
- le degré de croyance β_ℓ que la source soit fiable, sachant que la vérité est ω_ℓ , pour tout $\omega_\ell \notin A$.

La propriété de l'affaiblissement classique comme résultat d'une combinaison disjonctive, énoncée dans la remarque 4.4, trouve son équivalent dans l'affaiblissement contextuel par la proposition suivante.

Proposition 4.2. αm est égale à la combinaison disjonctive de m avec une fonction de masse m_0 définie par :

$$m_0(C) = \prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \bar{C}} \beta_\ell, \quad \forall C \subseteq \Omega. \quad (4.26)$$

Preuve : voir annexe A, paragraphe A.1. \square

Les deux propositions suivantes montrent que $\alpha \mathbf{G}$ de la proposition 4.1 et m_0 de la proposition 4.2 peuvent se décomposer en entités mathématiques simples, chacune correspondant à une opération d'affaiblissement contextuel élémentaire.

Proposition 4.3. La fonction de masse m_0 de la proposition 4.2 peut s'écrire de la manière suivante :

$$m_0 = m_1 \odot m_2 \odot \dots \odot m_K,$$

où chaque m_k est définie par :

$$\begin{aligned} m_k(\emptyset) &= \beta_k, \\ m_k(\{\omega_k\}) &= \alpha_k. \end{aligned}$$

Preuve : Immédiate à partir de l'équation 4.26 et de la définition de la combinaison disjonctive (équation 2.23). \square

Chaque fonction de masse m_k a la même forme que la fonction de masse m_0 de la proposition 4.2. Elles correspondent à un affaiblissement contextuel avec un vecteur d'affaiblissement α_k dont toutes les composantes sont nulles excepté la composante k égale à α_k . Ainsi, l'affaiblissement contextuel de vecteur d'affaiblissement $\alpha = (\alpha_1, \dots, \alpha_K)$ peut se décomposer en une séquence d'affaiblissements contextuels associés aux vecteurs d'affaiblissement $\alpha_1 = (\alpha_1, 0, \dots, 0)$, $\alpha_2 = (0, \alpha_2, 0, \dots, 0)$, \dots , $\alpha_K = (0, \dots, 0, \alpha_K)$:

$$\alpha m = \alpha_K(\alpha_{K-1}(\dots(\alpha_1 m)\dots)). \quad (4.27)$$

Corollaire 4.1. La matrice de généralisation $\alpha \mathbf{G}$, définie dans la proposition 4.1, peut s'écrire de la manière suivante :

$$\alpha \mathbf{G} = \prod_{k=1}^K \alpha_k \mathbf{G},$$

où $\alpha_k \mathbf{G}$ est la matrice de généralisation définie par :

$$\alpha_k G(A, B) = \begin{cases} 1 & \text{si } A = B \text{ et } \omega_k \in B, \\ \beta_k & \text{si } A = B \text{ et } \omega_k \notin B, \\ \alpha_k & \text{si } A \neq B \text{ et } A = \{\omega_k\} \cup B, \\ 0 & \text{sinon.} \end{cases} \quad (4.28)$$

Preuve : Immédiate à partir de la proposition 4.3. Chaque matrice de généralisation $\alpha_k \mathbf{G}$ correspond à un affaiblissement contextuel de vecteur d'affaiblissement α_k . \square

Une autre conséquence de la proposition 4.3 concerne la simplicité d'expression de l'opération d'affaiblissement contextuel en terme de fonctions d'implicabilité et de plausibilité.

Proposition 4.4. Soient αb et αpl les fonctions d'implicabilité et de plausibilité associées à αm . Ces fonctions peuvent être obtenues à partir de b et pl , les fonctions d'implicabilité et de plausibilité associées à m , de la manière suivante :

$$\alpha b(A) = b(A) \prod_{\omega_k \in \bar{A}} \beta_k, \quad \forall A \subseteq \Omega,$$

et

$$\alpha pl(A) = 1 - (1 - pl(A)) \prod_{\omega_k \in A} \beta_k, \quad \forall A \subseteq \Omega .$$

Preuve : Soit b_k la fonction d'implicabilité associée à m_k de la proposition 4.3, pour tout $k \in \llbracket 1, K \rrbracket$. On a :

$$b_k(A) = \begin{cases} \beta_k & \text{si } \omega_k \notin A \\ 1 & \text{sinon.} \end{cases}$$

À partir des propositions 4.2 et 4.3, on obtient $\alpha b = b \prod_{k=1}^K b_k$, d'où l'expression de αb . L'expression de αpl peut être aisément obtenue à partir de l'égalité $\alpha pl(A) = 1 - \alpha b(\bar{A})$.

□

4.4.3 Discussion

Le corollaire 4.1, et en particulier l'équation (4.28), mettent en lumière certains aspects de l'affaiblissement contextuel. Pour chaque k , chaque masse $m(B)$, $B \subseteq \Omega$, est inchangée si $\omega_k \in B$. En revanche, une fraction α_k de cette masse est transférée sur $\{\omega_k\} \cup B$ si $\omega_k \notin B$. Cela peut être interprété de la manière suivante : si la vérité est ω_k , le degré de croyance de l'agent dans le fait que la source est fiable est β_k ; en conséquence, cette fraction de la masse initialement allouée par la source à B reste sur B , tandis que la part restante est transférée à l'union de B et $\{\omega_k\}$. Cette opération est répétée pour chaque B et chaque k .

Notons que l'affaiblissement contextuel, tel que défini dans ce paragraphe (équation 4.24), *ne généralise pas* l'affaiblissement classique (équation 4.9) : comme nous le montrerons dans des exemples à venir, la solution obtenue par un affaiblissement de m avec des taux $\alpha_i = \alpha$, $i = 1, \dots, K$ est différente, en général, de la solution obtenue par un affaiblissement classique de m avec un unique taux α . En réalité, l'affaiblissement classique et cet affaiblissement contextuel sont deux cas particuliers d'un mécanisme d'affaiblissement plus général, qui sera introduit au paragraphe 4.5.

Nous pourrions aussi nous demander ce que cet affaiblissement deviendrait si l'expression des croyances sur la fiabilité de la source (équation 4.22) décrite au paragraphe 4.4.1 était modifiée. Une première alternative consiste à allouer α_k à $\{NR\}$, et à laisser la masse β_k sur $\{R\}$. La fonction de croyance de l'agent sur la fiabilité de la source est alors une masse Bayésienne. Une deuxième alternative consiste à allouer α_k à $\{NR\}$, et à transférer la masse β_k sur \mathcal{R} . Comme montré dans le paragraphe A.2 en annexe, le même résultat est obtenu dans le premier cas, tandis que la fonction de masse vide est obtenue dans le second cas.

4.4.4 Cas particuliers et exemples

Les matrices de généralisation ${}^\alpha G$ associées à l'opération d'affaiblissement contextuel dans les cas $K = 2$ et $K = 3$ sont illustrées dans les tableaux 4.1 et 4.2, respectivement.

Dans le cas $K = 2$, nous avons :

$$\begin{aligned} {}^\alpha m(\emptyset) &= \beta_1 \beta_2 m(\emptyset) \\ {}^\alpha m(\{\omega_1\}) &= \alpha_1 \beta_2 m(\emptyset) + \beta_2 m(\{\omega_1\}) \\ {}^\alpha m(\{\omega_2\}) &= \alpha_2 \beta_1 m(\emptyset) + \beta_1 m(\{\omega_1\}) \\ {}^\alpha m(\Omega) &= \alpha_1 \alpha_2 m(\emptyset) + \alpha_2 m(\{\omega_1\}) + \alpha_1 m(\{\omega_2\}) + m(\Omega) . \end{aligned}$$

Nous pouvons remarquer que dans le cas où $m(\emptyset) = 0$ et $\alpha_1 = \alpha_2$, l'affaiblissement contextuel est alors équivalent à l'affaiblissement classique. Ce n'est cependant pas le cas pour $K = 3$. En général, durant un affaiblissement contextuel, une fraction de chaque masse $m(A)$ est transférée à des sur-ensembles quelconques de A , et non uniquement à Ω , comme c'est le cas lors d'un affaiblissement classique.

Tableau 4.1 – Matrice de généralisation associée à l'affaiblissement contextuel dans le cas

$$K = 2.$$

$A \setminus B$	\emptyset	$\{\omega_1\}$	$\{\omega_2\}$	Ω
\emptyset	$\beta_1 \beta_2$			
$\{\omega_1\}$	$\alpha_1 \beta_2$	β_2		
$\{\omega_2\}$	$\beta_1 \alpha_2$		β_1	
Ω	$\alpha_1 \alpha_2$	α_2	α_1	1

Tableau 4.2 – Matrice de généralisation associée à l'affaiblissement contextuel dans le cas

$$K = 3.$$

$A \setminus B$	\emptyset	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\{\omega_1, \omega_3\}$	$\{\omega_2, \omega_3\}$	Ω
\emptyset	$\beta_1 \beta_2 \beta_3$							
$\{\omega_1\}$	$\alpha_1 \beta_2 \beta_3$	$\beta_2 \beta_3$						
$\{\omega_2\}$	$\beta_1 \alpha_2 \beta_3$		$\beta_1 \beta_3$					
$\{\omega_1, \omega_2\}$	$\alpha_1 \alpha_2 \beta_3$	$\alpha_2 \beta_3$	$\alpha_1 \beta_3$	β_3				
$\{\omega_3\}$	$\beta_1 \beta_2 \alpha_3$				$\beta_1 \beta_2$			
$\{\omega_1, \omega_3\}$	$\alpha_1 \beta_2 \alpha_3$	$\beta_2 \alpha_3$			$\alpha_1 \beta_2$	β_2		
$\{\omega_2, \omega_3\}$	$\beta_1 \alpha_2 \alpha_3$		$\beta_1 \alpha_3$		$\beta_1 \alpha_2$		β_1	
Ω	$\alpha_1 \alpha_2 \alpha_3$	$\alpha_2 \alpha_3$	$\alpha_1 \alpha_3$	α_3	$\alpha_1 \alpha_2$	α_2	α_1	1

Exemple 4.2 (Suite de l'exemple 4.1). *Supposons toujours que le capteur ait fourni l'information suivante au sujet d'une cible à identifier : $m(\{a\}) = 0.5$ et $m(\{r\}) =$*

0.5, et observons les résultats obtenus par affaiblissement contextuel à partir de différentes connaissances sur la fiabilité de ce capteur.

Cas 1 : La plausibilité α_1 que le capteur soit non fiable quand la cible est un avion est égale à 0.4 ; en revanche, le capteur est totalement fiable ($\alpha_2 = \alpha_3 = 0$) quand la cible s'avère être un hélicoptère ou un missile. Le vecteur d'affaiblissement est alors $\alpha_1 = (0.4, 0, 0)$. La matrice de généralisation $\alpha_1 G$ associée est la suivante :

$$\alpha_1 G = \begin{pmatrix} 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 1 \end{pmatrix} \begin{matrix} \emptyset \\ \{a\} \\ \{h\} \\ \{a, h\} \\ \{r\} \\ \{a, r\} \\ \{h, r\} \\ \Omega \end{matrix}$$

et la fonction de masse affaiblie $\alpha_1 m$ est donnée par :

$$\alpha_1 m(\{a\}) = 0.5, \quad \alpha_1 m(\{r\}) = 0.5 \times 0.6 = 0.3, \quad \alpha_1 m(\{a, r\}) = 0.5 \times 0.4 = 0.2 .$$

Nous pouvons remarquer qu'une fraction 0.4 de la masse initialement allouée à $\{r\}$ a été transférée à $\{a, r\}$, ce qui peut être interprété de la manière suivante : si la cible est un missile ou un hélicoptère, le capteur est totalement fiable et il détectera de manière adéquate la cible. En revanche, si la cible est un avion, alors le capteur n'est pas fiable, et il peut d'une manière erronée déclarer la cible comme étant un missile ou un hélicoptère. En conséquence, quand le capteur signale un missile, en réalité il peut s'agir d'un missile ou d'un objet mal identifié par le capteur, dans ce cas un avion. Si la vérité avait été un hélicoptère, comme le capteur est totalement fiable dans la détection des hélicoptères, le capteur aurait décidé hélicoptère et non un missile.

Cas 2 : La plausibilité α_2 que le capteur soit non fiable quand la cible est un hélicoptère est égale à 0.6 ; en revanche le capteur est totalement fiable ($\alpha_1 = \alpha_3 = 0$) quand la cible est un avion ou un missile. Le vecteur d'affaiblissement correspondant est $\alpha_2 = (0, 0.6, 0)$. La matrice de généralisation $\alpha_2 G$ associée est :

$$\alpha_2 G = \begin{pmatrix} 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 1 \end{pmatrix} \begin{matrix} \emptyset \\ \{a\} \\ \{h\} \\ \{a, h\} \\ \{r\} \\ \{a, r\} \\ \{h, r\} \\ \Omega \end{matrix}$$

et la fonction de masse affaiblie $\alpha_2 m$ est donnée par

$$\begin{aligned} \alpha_2 m(\{a\}) &= 0.5 \times 0.4 = 0.2, & \alpha_2 m(\{a, h\}) &= 0.5 \times 0.6 = 0.3, \\ \alpha_2 m(\{r\}) &= 0.5 \times 0.4 = 0.2, & \alpha_2 m(\{h, r\}) &= 0.5 \times 0.6 = 0.3 . \end{aligned}$$

L'interprétation est similaire à celle donnée dans le cas 1 : cette fois, les masses données à $\{a\}$ et $\{r\}$ sont partiellement transférées, respectivement, à $\{a, h\}$ et $\{h, r\}$, afin de prendre en compte la faible reconnaissance des hélicoptères.

Cas 3 : Nous avons à la fois $\alpha_1 = 0.4$ et $\alpha_2 = 0.6$, signifiant que le capteur a une plausibilité 0.4 de ne pas être fiable quand la cible est un avion, et 0.6 quand celle-ci est un hélicoptère. Comme précédemment, le capteur est supposé entièrement fiable quand la cible est un missile. Le vecteur d'affaiblissement est $\alpha = (0.4, 0.6, 0)$. La matrice de généralisation ${}^\alpha G$ est alors :

$${}^\alpha G = \begin{pmatrix} 0.24 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.16 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.36 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0.24 & 0.6 & 0.4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.24 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.16 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.36 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.24 & 0.6 & 0.4 & 1 \end{pmatrix} \begin{matrix} \emptyset \\ \{a\} \\ \{h\} \\ \{a, h\} \\ \{r\} \\ \{a, r\} \\ \{h, r\} \\ \Omega \end{matrix}$$

et ${}^\alpha m$ est donnée par

$$\begin{aligned} {}^\alpha m(\{a\}) &= 0.2, & {}^\alpha m(\{a, h\}) &= 0.3, \\ {}^\alpha m(\{r\}) &= 0.12, & {}^\alpha m(\{a, r\}) &= 0.08, \\ {}^\alpha m(\{h, r\}) &= 0.18, & {}^\alpha m(\Omega) &= 0.12. \end{aligned}$$

Remarquons qu'en conséquence de la proposition 4.3, nous avons

$${}^\alpha G = \alpha_1 G \alpha_2 G = \alpha_2 G \alpha_1 G,$$

et ${}^\alpha m = \alpha_1({}^{\alpha_2} m) = \alpha_2({}^{\alpha_1} m)$. L'affaiblissement contextuel de vecteur d'affaiblissement $\alpha = (0.4, 0.6, 0)$ est ainsi équivalent à un affaiblissement contextuel de vecteur d'affaiblissement $\alpha_1 = (0.4, 0, 0)$ comme dans le cas 1, suivi d'un affaiblissement de vecteur d'affaiblissement $\alpha_2 = (0, 0.6, 0)$ comme dans le cas 2. Par exemple la masse ${}^\alpha m(\Omega) = 0.12$ peut être expliquée par le transfert de 40 % de la masse $m(\{r\}) = 0.5$ sur $\{a, r\}$, de laquelle 60% ont été transférés sur Ω . Quand le capteur signale un missile, en réalité il peut s'agir d'un missile ou d'un objet mal identifié par le capteur ; dans ce cas un avion ou un hélicoptère. Nous pouvons remarquer que lors du transfert d'une masse allouée à un sous-ensemble B de Ω , les fiabilités sur les éléments contenus dans B n'interviennent pas, seules les fiabilités des éléments appartenant au complémentaire de B dans Ω interviennent (figure 4.5).

Exemple 4.3 (Inspiré de l'exemple de Zadeh [125]). Soient $m_1, m_2 \in \mathcal{M}^\Omega$, $\Omega = \{a, b, c\}$, issues de deux capteurs C_1 et C_2 . Supposons, d'un côté que $m_1(\{a\}) = 0.9$ et $m_1(\{c\}) = 0.1$, et de l'autre que $m_2(\{b\}) = 0.9$ et $m_2(\{c\}) = 0.1$. Une combinaison conjonctive de ces fonctions apportent une forte masse sur le vide : $m_1 \odot_2(\{c\}) = 0.01$ et $m_1 \odot_2(\emptyset) = 0.99$.

Supposons que l'information supplémentaire suivante soit disponible : « le capteur C_1 ne reconnaît pas les éléments b de Ω ». Celle-ci peut être prise en compte via

un affaiblissement contextuel de la fonction de masse m_1 fournie par le capteur C_1 avec un vecteur d'affaiblissement $\alpha = (\alpha_a = 0, \alpha_b = 1, \alpha_c = 0)$. Nous avons alors $\alpha m_1(\{a, b\}) = 0.9$ et $\alpha m_1(\{b, c\}) = 0.1$, et la combinaison conjonctive des deux capteurs C_1 et C_2 avec cette nouvelle information conduit à $\alpha m_1 \odot m_2(\{b\}) = 0.9$, $\alpha m_1 \odot m_2(\{c\}) = 0.01$, et $\alpha m_1 \odot m_2(\emptyset) = 0.09$. Ainsi, sur cet exemple, la nouvelle information a permis d'expliquer une partie du conflit entre m_1 et m_2 , sans pour autant affaiblir totalement l'information issue du capteur C_1 .

Pour conclure sur cet affaiblissement, retenons que lors d'un affaiblissement contextuel, chaque masse initialement allouée à un sous-ensemble B de Ω est transférée à tout sur-ensemble A de B , proportionnellement à la non-fiabilité de la source dans la détection des éléments de $A \setminus B$, et à la fiabilité de la source dans la détection des éléments de \bar{A} . Ceci est illustré sur la figure 4.5, qui est une représentation graphique de l'équation (4.25) de $\alpha G(A, B)$.

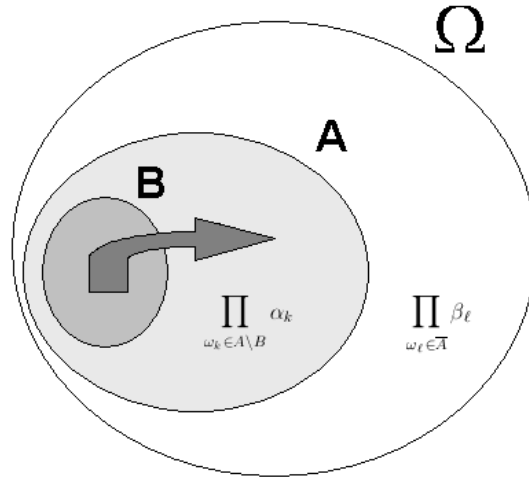


Figure 4.5 – Illustration d'un transfert de masse lors d'un affaiblissement contextuel.

4.5 Affaiblissement contextuel basé sur un grossissement

Dans le paragraphe précédent, nous avons supposé que la fiabilité de la source d'information pouvait être exprimée conditionnellement à chaque élément ω_k de Ω . Cependant, dans certaines situations, des informations aussi détaillées ne sont pas disponibles. Néanmoins, il est toujours possible d'exprimer la fiabilité de la source conditionnellement à des regroupements d'éléments de Ω , c'est-à-dire à chaque élément d'un grossissement de Ω . Dans ce paragraphe, le modèle d'affaiblissement basé sur cette hypothèse est développé. Celui-ci généralise l'affaiblissement contextuel obtenu dans le paragraphe précédent, ainsi que l'affaiblissement classique.

4.5.1 Expression de la fiabilité

Soient $\Theta = \{\theta_1, \dots, \theta_L\}$ un grossissement de Ω , $\theta_1, \dots, \theta_L$ forment une partition de Ω , et $m_{Ag}^{\mathcal{R}}[\theta_k]$ une fonction de masse définie sur \mathcal{R} quantifiant la croyance de l'agent Ag en la fiabilité de la source, quand $x \in \theta_k$. L'élément θ_k constitue alors un contexte plus général que ceux considérés dans le paragraphe précédent. Supposons que chaque $m_{Ag}^{\mathcal{R}}[\theta_k]$, $k = 1, \dots, L$ ait la forme suivante :

$$\begin{cases} m_{Ag}^{\mathcal{R}}[\theta_k](\{R\}) &= \beta_k, \\ m_{Ag}^{\mathcal{R}}[\theta_k](\mathcal{R}) &= \alpha_k, \end{cases} \quad (4.29)$$

où, comme précédemment, $\beta_k = 1 - \alpha_k$.

Comme au paragraphe 4.4.1, nous voulons combiner ces L fonctions de croyance conditionnelles avec $m_{Ag}^{\Omega}[R]$. Cela peut être réalisé en calculant tous les déconditionnements, en les combinant conjonctivement, et en projetant le résultat sur Ω :

$$m_{Ag}^{\Omega} [m_S^{\Omega}, m_{Ag}^{\mathcal{R}}[\theta_1], \dots, m_{Ag}^{\mathcal{R}}[\theta_L]] = (m_{Ag}^{\Omega}[R]^{\uparrow\Omega \times \mathcal{R}} \odot m_{Ag}^{\mathcal{R}}[\theta_1]^{\uparrow\Omega \times \mathcal{R}} \odot \dots \odot m_{Ag}^{\mathcal{R}}[\theta_L]^{\uparrow\Omega \times \mathcal{R}})^{\downarrow\Omega}. \quad (4.30)$$

Le résultat de cette combinaison ne dépend que de m_S^{Ω} , Θ , et du vecteur $\alpha = (\alpha_1, \dots, \alpha_L)$ composé des taux d'affaiblissement. Ainsi, cet affaiblissement sera noté $\mathfrak{G}m_S^{\Omega}$, ou plus simplement $\mathfrak{G}m$. La transformation de m en $\mathfrak{G}m$ sera appelée *affaiblissement Θ -contextuel* de vecteur d'affaiblissement α . Notons que l'affaiblissement contextuel défini au paragraphe 4.4 correspond au cas particulier $L = K$ et $\theta_k = \{\omega_k\}$, $k = 1, \dots, K$. Ainsi, l'affaiblissement contextuel est équivalent à l'affaiblissement Ω -contextuel. Comme nous le verrons un peu plus loin, l'affaiblissement classique est égale à l'affaiblissement $\{\Omega\}$ -contextuel, c'est-à-dire au cas $\Theta = \{\Omega\}$.

4.5.2 Expression de $\mathfrak{G}m$

Pour tout $A \subseteq \Omega$, on définit :

$$A_* = \bigcup_{\{\theta \in \Theta, \theta \subseteq A\}} \theta,$$

$$A^* = \bigcup_{\{\theta \in \Theta, \theta \cap A \neq \emptyset\}} \theta,$$

et

$$\mathcal{C} = \{A \subseteq \Omega \mid \exists I \subseteq \{1, \dots, L\}, A = \bigcup_{i \in I} \theta_i\}.$$

Les ensembles A_* et A^* , illustrés sur la figure 4.6, sont respectivement, le plus grand élément de \mathcal{C} inclus dans A , et le plus petit élément de \mathcal{C} contenant A , et sont respectivement appelés réductions intérieure et extérieure de A .

Avec ces notations, les propositions suivantes sont vérifiées :

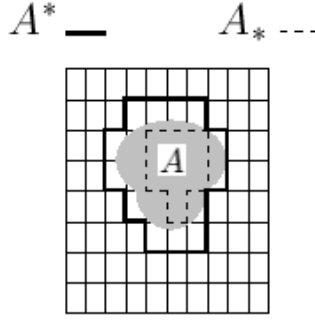


Figure 4.6 – Illustration des réductions extérieure et intérieure d'un sous-ensemble A de Ω avec un grossissement « en pavé » de Ω .

Proposition 4.5. Soient $m \in \mathcal{M}^\Omega$, $\Theta = \{\theta_1, \dots, \theta_L\}$ un grossissement de Ω , et $\alpha = (\alpha_1, \dots, \alpha_L) \in [0, 1]^L$. La fonction de croyance $\mathfrak{G}m$, résultant de l'affaiblissement Θ -contextuel de m avec le vecteur d'affaiblissement α , est donnée par :

$$\mathfrak{G}m(A) = \sum_{B \subseteq A} \mathfrak{G}G(A, B)m(B), \quad \forall A \subseteq \Omega, \quad (4.31)$$

avec :

$$\mathfrak{G}G(A, B) = \begin{cases} \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup \theta_\ell = \bar{A}^*} \beta_\ell & \text{si } \exists C \in \mathcal{C}, B \cup C = A, \\ 0 & \text{sinon.} \end{cases} \quad (4.32)$$

Preuve : Voir annexe A, paragraphe A.3. \square

Proposition 4.6. La fonction de masse $\mathfrak{G}m$ est le résultat de la combinaison disjunctive de m avec $m_0 \in \mathcal{M}^\Omega$ définie par

$$m_0(C) = \begin{cases} \prod_{\cup \theta_k = C} \alpha_k \prod_{\cup \theta_\ell = \bar{C}} \beta_\ell & \text{si } C \in \mathcal{C}, \\ 0 & \text{sinon.} \end{cases} \quad (4.33)$$

Preuve : Voir annexe A, paragraphe A.3. \square

Comme dans le cas de l'affaiblissement Ω -contextuel, la fonction de masse m_0 introduite dans la proposition 4.6, ainsi que la matrice de généralisation $\mathfrak{G}G$ introduite dans la proposition 4.5 admettent une décomposition simple, décrite dans les deux propositions suivantes.

Proposition 4.7. La fonction de masse m_0 définie dans la proposition 4.6 peut être décomposée de la manière suivante :

$$m_0 = m_1 \circledast m_2 \circledast \dots \circledast m_L,$$

où, pour tout $\ell \in [1, L]$, m_ℓ est définie par

$$\begin{aligned} m_\ell(\emptyset) &= \beta_\ell, \\ m_\ell(\theta_\ell) &= \alpha_\ell. \end{aligned}$$

Preuve : Immédiate à partir de l'équation 4.33 et de la définition de la combinaison disjonctive (équation 2.23). \square

Chaque fonction de masse m_ℓ correspond à un affaiblissement Θ -contextuel associé à un vecteur d'affaiblissement α_ℓ dont toutes les composantes sont égales à 0, sauf la composante ℓ égale à α_ℓ . Ainsi, l'affaiblissement Θ -contextuel de vecteur d'affaiblissement $\alpha = (\alpha_1, \dots, \alpha_L)$ peut se décomposer en L affaiblissements Θ -contextuels de vecteurs d'affaiblissement respectifs $\alpha_1, \dots, \alpha_L$.

Corollaire 4.2. *La matrice de généralisation $\mathfrak{G}\mathbf{G}$, définie dans la proposition 4.5, peut s'exprimer de la manière suivante :*

$$\mathfrak{G}\mathbf{G} = \prod_{\ell=1}^L \mathfrak{G}^{\alpha_\ell} \mathbf{G},$$

où chaque matrice de généralisation $\mathfrak{G}^{\alpha_\ell} \mathbf{G}$ est associée à la fonction de masse m_ℓ^Ω définie par :

$$\mathfrak{G}^{\alpha_\ell} G(A, B) = \begin{cases} 1 & \text{si } A = B \text{ et } \theta_\ell \subseteq B, \\ \beta_\ell & \text{si } A = B \text{ et } \theta_\ell \not\subseteq B, \\ \alpha_\ell & \text{si } A \neq B \text{ et } A = \theta_\ell \cup B, \\ 0 & \text{sinon.} \end{cases} \quad (4.34)$$

Preuve : Immédiate à partir de la proposition 4.7. Chaque matrice de généralisation $\mathfrak{G}^{\alpha_\ell} \mathbf{G}$ correspond à la composition disjonctive avec m_ℓ^Ω , et à l'affaiblissement Θ -contextuel de vecteur d'affaiblissement α_ℓ . \square

Proposition 4.8. *Soient $\mathfrak{G}b$ et $\mathfrak{G}pl$ les fonctions d'implicabilité et de plausibilité associées à $\mathfrak{G}m$. Celles-ci peuvent être obtenues, respectivement, à partir de b et pl , les fonctions d'implicabilité et de plausibilité associées à m .*

$$\mathfrak{G}b(A) = b(A) \prod_{\cup \theta_\ell = \overline{A}^*} \beta_\ell, \quad \forall A \subseteq \Omega \quad (4.35)$$

et

$$\mathfrak{G}pl(A) = 1 - (1 - pl(A)) \prod_{\cup \theta_\ell = A^*} \beta_\ell, \quad \forall A \subseteq \Omega. \quad (4.36)$$

Preuve : Soit b_ℓ la fonction d'implicabilité associée à m_ℓ définie à la proposition 4.7, pour tout $\ell \in [1, L]$. On a

$$b_\ell(A) = \begin{cases} \beta_\ell & \text{si } \theta_\ell \not\subseteq A \\ 1 & \text{sinon.} \end{cases}$$

Ainsi, à partir des propositions 4.6 et 4.7, $\mathfrak{G}b = b \prod_{\ell=1}^L b_\ell$, d'où l'expression obtenue pour $\mathfrak{G}b$. L'expression de $\mathfrak{G}pl$ est aisément obtenue à partir de l'égalité $\mathfrak{G}pl(A) = 1 - \mathfrak{G}b(\overline{A})$. \square

Comme mentionné précédemment, l'affaiblissement Θ -contextuel généralise l'affaiblissement contextuel défini au paragraphe 4.4. Il généralise aussi l'affaiblissement classique introduit par Shafer et rappelé au paragraphe 4.3, comme le montre la proposition suivante.

Proposition 4.9. *Si $\Theta = \{\Omega\}$, autrement dit, si Θ correspond à la partition triviale de Ω composée uniquement d'une classe. Alors, l'affaiblissement Θ -contextuel, associé à un vecteur d'affaiblissement à une composante, est identique à l'affaiblissement classique de taux d'affaiblissement cette composante.*

En remplaçant Θ par $\{\Omega\}$, il est possible d'obtenir une preuve de cette proposition à partir de chaque proposition de ce paragraphe. En voici une des plus courtes.

Preuve : D'après la proposition 4.6, ${}^\alpha m$ est le résultat de la combinaison disjonctive de m avec m_0 définie par $m_0(\emptyset) = 1 - \alpha$ et $m_0(\Omega) = \alpha$. Ainsi, par la remarque 4.4, ${}^\alpha m$ est égale à l'affaiblissement classique de m . \square

Exemple 4.4 (Retour sur l'exemple 4.1). *Supposons que l'agent ait un degré de croyance égal à 0.4 dans le fait que la source soit fiable quand la vérité est un avion, et, un degré de croyance égal à 0.9 dans le fait que la source soit fiable quand la vérité est un hélicoptère ou un missile. Le grossissement associé est $\Theta = \{\theta_1, \theta_2\}$, où $\theta_1 = \{a\}$ et $\theta_2 = \{h, r\}$. Soit $\alpha = (\alpha_1, \alpha_2)$. La matrice de généralisation associée est illustrée dans le tableau 4.3, la notation $\Omega = \{\omega_1, \omega_2, \omega_3\}$ est employée afin de faciliter les comparaisons avec le tableau 4.2.*

Nous avons $\alpha_1 = 0.6$ et $\alpha_2 = 0.1$. Si, comme précédemment, la fonction de croyance fournie par le capteur est définie par $m(\{a\}) = 0.5$ et $m(\{r\}) = 0.5$, le résultat de l'affaiblissement Θ -contextuel de m est :

$$\begin{aligned} \mathfrak{G}m(\{a\}) &= 0.45, & \mathfrak{G}m(\Omega) &= 0.05 + 0.03 = 0.08, \\ \mathfrak{G}m(\{r\}) &= 0.18, & \mathfrak{G}m(\{a, r\}) &= 0.27, & \mathfrak{G}m(\{h, r\}) &= 0.02. \end{aligned}$$

Tableau 4.3 – Matrice de généralisation associée à l'affaiblissement Θ -contextuel dans le cas $K = 3$ et $\Theta = \{\{\omega_1\}, \{\omega_2, \omega_3\}\}$.

	\emptyset	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\{\omega_1, \omega_3\}$	$\{\omega_2, \omega_3\}$	Ω
\emptyset	$\beta_1\beta_2$							
$\{\omega_1\}$	$\alpha_1\beta_2$	β_2						
$\{\omega_2\}$			$\beta_1\beta_2$					
$\{\omega_1, \omega_2\}$			$\alpha_1\beta_2$	β_2				
$\{\omega_3\}$					$\beta_1\beta_2$			
$\{\omega_1, \omega_3\}$					$\alpha_1\beta_2$	β_2		
$\{\omega_2, \omega_3\}$	$\beta_1\alpha_2$		$\beta_1\alpha_2$		$\beta_1\alpha_2$		β_1	
Ω	$\alpha_1\alpha_2$	α_2	$\alpha_1\alpha_2$	α_2	$\alpha_1\alpha_2$	α_2	α_1	1

4.6 Apprentissage automatique des coefficients d'affaiblissement

Dans la pratique, le vecteur d'affaiblissement d'un affaiblissement contextuel, peut être soit fixé par un expert, soit appris à partir de données. Cette dernière

approche est considérée dans ce paragraphe. Nous généralisons une méthode d'apprentissage automatique des taux d'affaiblissement de l'affaiblissement classique introduite par Elouedi, Mellouli et Smets dans [40].

4.6.1 Apprentissage de la fiabilité d'un capteur seul

De même que dans [40], supposons que nous disposions d'un ensemble d'apprentissage composé de n objets o_1, \dots, o_n . Chaque objet o_i appartient à une classe dans $\Omega = \{\omega_1, \dots, \omega_K\}$. La classe de l'objet o_i est encodée par K variables binaires $\delta_{i,k}$, telles que $\delta_{i,k} = 1$ si la classe de o_i est ω_k , et $\delta_{i,k} = 0$ sinon.

Supposons qu'une source d'information S , un capteur ou un expert, fournisse, pour chaque objet o_i , une fonction de masse $m_S^\Omega\{o_i\}$, quantifiant ses croyances sur la véritable classe de o_i . En s'inspirant de travaux déjà menés en reconnaissance de formes par Zouhal et Denœux [128], les auteurs de [40] proposent de trouver le taux d'affaiblissement α minimisant la mesure de divergence suivante entre les croyances et les observations :

$$E_{bet}(\alpha) = \sum_{i=1}^n \sum_{k=1}^K (\alpha \text{Bet} P^\Omega\{o_i\}(\omega_k) - \delta_{i,k})^2. \quad (4.37)$$

La minimisation de $E_{bet}(\alpha)$ est un problème non linéaire sous contraintes ($0 \leq \alpha \leq 1$). Quand toutes les fonctions de masse sont normalisées, c'est-à-dire lorsque $m^\Omega\{o_i\}(\emptyset) = 0$ pour tout i , $E_{bet}(\alpha)$ est une fonction quadratique de α , dont le minimum peut être analytiquement obtenu [40].

Cette méthode d'apprentissage automatique peut être aisément généralisée à l'affaiblissement contextuel. Dans ce paragraphe nous considérons uniquement l'affaiblissement Ω -contextuel, mais cette approche s'applique tout aussi bien à l'affaiblissement Θ -contextuel, une fois le grossissement Θ choisi. En gardant le même critère d'erreur $E_{bet}(\boldsymbol{\alpha})$, définie par l'équation 4.37, où l'affaiblissement contextuel de vecteur d'affaiblissement $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ est utilisé à la place de l'affaiblissement classique, le problème d'optimisation est cette fois un problème de minimisation non linéaire d'une fonction de K variables, sous les contraintes $0 \leq \alpha_k \leq 1$, pour tout $k \in \llbracket 1, K \rrbracket$. Celui-ci peut être résolu par une procédure de minimisation non linéaire avec contraintes.

Cependant, le problème est d'une plus faible complexité si nous remplaçons $E_{bet}(\boldsymbol{\alpha})$ par la fonction d'erreur alternative suivante :

$$E_{pl}(\boldsymbol{\alpha}) = \sum_{i=1}^n \sum_{k=1}^K (\alpha p l^\Omega\{o_i\}(\{\omega_k\}) - \delta_{i,k})^2, \quad (4.38)$$

qui semble aussi raisonnable que $E_{bet}(\boldsymbol{\alpha})$ (4.37). En effet, par la proposition 4.4, $\alpha p l^\Omega\{o_i\}(\{\omega_k\})$ est une fonction affine de α_k :

$$\begin{aligned} \alpha p l^\Omega\{o_i\}(\{\omega_k\}) &= 1 - (1 - p l^\Omega\{o_i\}(\{\omega_k\}))\beta_k \\ &= \alpha_k(1 - p l^\Omega\{o_i\}(\{\omega_k\})) + p l^\Omega\{o_i\}(\{\omega_k\}). \end{aligned}$$

Notons :

- $\mathbf{pl}_i = (pl^\Omega\{o_i\}(\{\omega_1\}), \dots, pl^\Omega\{o_i\}(\{\omega_K\}))^T$ le vecteur de dimension K composé des plausibilités des singletons pour l'objet o_i ;
- $\text{diag}(1 - \mathbf{pl}_i)$ la matrice diagonale de dimension $K \times K$ dont les éléments diagonaux sont les compléments à un des composantes de \mathbf{pl}_i ; et
- $\boldsymbol{\delta}_i = (\delta_{i,1}, \dots, \delta_{i,K})^T$ le vecteur de dimension K composé des indicateurs de vérité pour l'objet o_i .

En posant

$$\mathbf{Q} = \begin{bmatrix} \text{diag}(1 - \mathbf{pl}_1) \\ \vdots \\ \text{diag}(1 - \mathbf{pl}_n) \end{bmatrix}, \text{ et } \mathbf{d} = \begin{bmatrix} \boldsymbol{\delta}_1 - \mathbf{pl}_1 \\ \vdots \\ \boldsymbol{\delta}_n - \mathbf{pl}_n \end{bmatrix},$$

l'équation (4.38) peut alors s'écrire sous la forme matricielle suivante :

$$E_{pl}(\boldsymbol{\alpha}) = \|\mathbf{Q}\boldsymbol{\alpha} - \mathbf{d}\|^2. \quad (4.39)$$

Il est alors clair que la minimisation de E_{pl} est un problème quadratique sous contraintes, qui peut être résolu efficacement en utilisant des algorithmes standards itératifs.

Exemple 4.5. *Considérons le même problème de reconnaissance de cible introduit dans l'exemple 4.1, ainsi que les données du tableau 4.4 issues de [40]. Quatre objets dont la classe est connue ont été classés par deux capteurs S_1 et S_2 . Chaque capteur a fourni, pour chaque objet, une fonction de masse définie sur Ω illustrée dans le tableau 4.4.*

Dans cet exemple, ces données sont utilisées afin d'optimiser les taux d'affaiblissement à appliquer à chaque capteur individuellement. D'après [40], l'apprentissage d'un taux d'affaiblissement par la minimisation de (4.37) conduit à $\alpha = 0.66$ pour le capteur S_1 , et $\alpha = 0.52$ pour le capteur S_2 . Avec notre approche, c'est-à-dire par apprentissage d'un vecteur composé de trois taux d'affaiblissement en minimisant (4.39), nous obtenons $\boldsymbol{\alpha} = (0.24, 0, 0)$ pour le capteur S_1 , et $\boldsymbol{\alpha} = (0.26, 0, 0)$ pour le capteur S_2 . Avec un ensemble d'apprentissage de plus grande taille, ces résultats indiqueraient que ces deux capteurs sont moins fiables dans le cas où la vérité est un avion, et que le capteur S_1 est légèrement meilleur que S_2 dans cette détection. Par rapport à l'apprentissage d'un taux d'affaiblissement seul, une information plus fine a pu être apprise : ce ne sont pas les hélicoptères ni les missiles qui posent problème aux capteurs. Bien entendu, ces interprétations doivent être relativisées du fait d'un si petit ensemble d'apprentissage, qui ne joue ici qu'un rôle d'illustration.

Remarque 4.6 (Concernant le choix entre E_{bet} et E_{pl}). *Aucun des deux critères d'erreur E_{bet} (4.37) et E_{pl} (4.38) ne semble mieux justifié que l'autre. Cependant, la conversion d'une relation complexe entre une fonction de croyance et la vérité en un simple nombre apporte obligatoirement une perte d'information. Le choix de E_{bet} , basée sur une mesure de probabilité, affecte la même erreur à une fonction de masse Bayésienne et à l'ignorance totale. Le choix de E_{pl} alloue une même erreur au cas où une plausibilité nulle est donnée à la vraie classe, et au cas où une plausibilité de 1 est donnée à une mauvaise classe. Or la première erreur semble plus grave que la seconde. Tout en conservant des qualités de complexité adéquates, il serait peut-être intéressant d'étudier d'autres critères d'erreur dans de futurs travaux.*

Tableau 4.4 – Données de l'exemple 4.5 issues de [40].

	a	h	r	$\{a, h\}$	$\{a, r\}$	$\{h, r\}$	Ω	vérité
$m_{S_1}\{o_1\}$	0	0	0.5	0	0	0.3	0.2	a
$m_{S_1}\{o_2\}$	0	0.5	0.2	0	0	0	0.3	h
$m_{S_1}\{o_3\}$	0	0.4	0	0	0.6	0	0	a
$m_{S_1}\{o_4\}$	0	0	0	0	0.6	0.4	0	r
$m_{S_2}\{o_1\}$	0	0	0	0.7	0	0	0.3	a
$m_{S_2}\{o_2\}$	0.3	0	0	0.4	0	0	0.3	h
$m_{S_2}\{o_3\}$	0.2	0	0	0	0	0.6	0.2	a
$m_{S_2}\{o_4\}$	0	0	0	0	0	1	0	r

Remarque 4.7. Dans ce paragraphe, seul le cas de l'affaiblissement Ω -contextuel a été traité. La généralisation à l'affaiblissement Θ -contextuel est immédiate une fois le grossissement Θ fixé. Cette contrainte ne semble pas rédhibitoire, car le choix de Θ est généralement guidé par les informations disponibles relatives à la fiabilité de la source, qui sont souvent intrinsèques à l'application traitée. Cependant, si ces informations n'existent pas, nous pourrions envisager de chercher le grossissement optimal Θ , ce qui constituerait un autre problème. Si la cardinalité de Ω est petite, une recherche exhaustive de la partition optimale peut être envisagée. Le développement de solutions plus efficaces pour traiter ce problème est laissé en perspective.

4.6.2 Apprentissage de la fiabilité d'un groupe de capteurs

Quand plusieurs capteurs fournissant des informations distinctes sont disponibles, une stratégie usuelle de combinaison consiste à d'abord corriger les informations fournies par chaque capteur par un affaiblissement, puis à combiner conjonctivement ces résultats. Dans ce cas, plutôt que d'optimiser individuellement les capteurs (figure 4.7(a)), il est préférable d'optimiser directement les performances de la combinaison de ce groupes de capteurs (figure 4.7(b)).

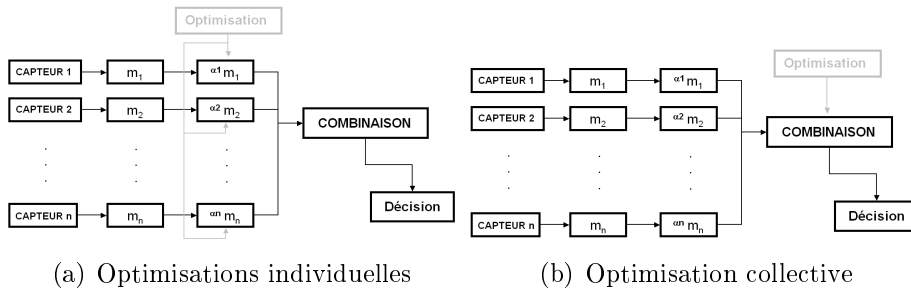


Figure 4.7 – Optimisation des coefficients d'affaiblissement lors d'une combinaison.

Pour simplifier, supposons disposer de deux capteurs S_1 et S_2 . Soient $\alpha_1 pl_{S_1}^\Omega \{o_i\}$ la plausibilité fournie par le capteur affaibli avec un vecteur d'affaiblissement α_1 , et $\alpha_2 pl_{S_2}^\Omega \{o_i\}$ la plausibilité fournie par le capteur S_2 affaibli à l'aide du vecteur α_2 . La

plausibilité d'un singleton étant égale à sa communalité, la plausibilité d'un singleton après combinaison conjonctive est égale au produit des plausibilités affaiblies issues de ces deux sources. Pour tout $k \in \llbracket 1, K \rrbracket$, la plausibilité de $\{\omega_k\}$ est égale à

$$\alpha^1 pl_{S_1}^\Omega \{o_i\}(\{\omega_k\}) \times \alpha^2 pl_{S_2}^\Omega \{o_i\}(\{\omega_k\}).$$

Les vecteurs d'affaiblissement α_1 et α_2 peuvent ainsi être appris pour minimiser le critère d'erreur suivant :

$$E_{pl}(\alpha_1, \alpha_2) = \sum_{i=1}^n \sum_{k=1}^K (\alpha^1 pl_{S_1}^\Omega \{o_i\}(\{\omega_k\}) \times \alpha^2 pl_{S_2}^\Omega \{o_i\}(\{\omega_k\}) - \delta_{i,k})^2. \quad (4.40)$$

Notons que ce critère n'est plus quadratique. Il peut être minimisé par l'emploi d'une procédure générale d'optimisation non linéaire sous contraintes.

Exemple 4.6. Avec les données du tableau 4.4, nous obtenons $\alpha_1 = (0.45, 0, 0)$ et $\alpha_2 = (0.39, 1, 0)$. Notons que, ce résultat ne peut être strictement comparé avec le résultat obtenu dans l'exemple 4.5 : dans le premier cas, nous optimisons les performances individuelles de chaque capteur, alors que dans cet exemple, nous optimisons les performances de la combinaison de ces deux capteurs. Avec un ensemble d'apprentissage plus grand, ces résultats indiqueraient que dans cette combinaison, le capteur S_2 n'est pas exploité pour détecter les hélicoptères. La seconde composante du vecteur α_2 indiquant une absence totale de fiabilité si la vérité est un hélicoptère, l'élément singleton « hélicoptère » sera présent dans chaque élément focal de la masse affaiblie en provenance de S_2 . En revanche, les informations de reconnaissance d'hélicoptères et de missiles issues du capteur S_1 seront pleinement prise en compte. Au final, l'apprentissage de ces vecteurs d'affaiblissement est équivalent à l'apprentissage d'une stratégie de combinaison de ces deux capteurs.

4.7 Mécanismes généraux de correction

Comme nous l'avons évoqué en introduction de ce chapitre, l'opération d'affaiblissement n'est pas la seule opération de correction d'une fonction de croyance. Dans ce paragraphe, nous rappelons les mécanismes déjà introduits permettant de renforcer ou de contredire une fonction de croyance, et présentons une nouvelle interprétation de ces mécanismes ainsi qu'une structure générale de correction de fonctions de croyance incluant comme cas particulier tous ces mécanismes.

4.7.1 Renforcement de Dencœux et Smets [29]

L'opération d'affaiblissement permet d'atténuer l'information fournie par une source. Or, il peut arriver qu'une source d'information soit trop prudente, et dans ce cas, contrairement au résultat d'un affaiblissement, nous souhaiterions « renforcer » l'information fournie par une telle source.

Dans [29], Dencœux et Smets ont introduit la notion de *renforcement*¹ d'une fonction de croyance. À l'origine, ce processus consiste à inverser le processus d'affaiblissement.

¹Pour l'instant, seul le terme anglais « de-discounting » a été employé.

Formellement, on suppose un agent Ag recevant une fonction de masse ${}^\alpha m$, affaiblie par un taux $\alpha < 1$. Si Ag connaît α , il peut recalculer m en inversant l'opération d'affaiblissement (équation 4.10), ainsi :

$$m = \frac{{}^\alpha m - \alpha m_\Omega}{1 - \alpha}. \quad (4.41)$$

Sinon, Ag peut imaginer n'importe quelle valeur de α dans l'intervalle $[0, {}^\alpha m(\Omega)]$. En effet, comme montré dans [29], ${}^\alpha m(\Omega)$ est la plus grande valeur de α telle que la fonction m obtenue par l'équation 4.41 demeure une fonction de masse.

L'opération décrite par l'équation (4.41), avec $\alpha \in [0, {}^\alpha m(\Omega)]$, est appelée *opération de renforcement* de ${}^\alpha m$. Durant cette transformation, une fraction de la masse allouée à Ω est uniformément distribuée à chaque élément focal, contrairement à un affaiblissement, où une fraction de la masse de chaque élément focal est transférée sur Ω . Ainsi, ce processus transforme une fonction de croyance en une fonction de croyance plus informative, donc renforcée.

Renforcer une fonction de masse m non vide avec la plus grande valeur possible pour α conduit à la *fonction de masse totalement renforcée* de m , notée ${}^{tr}m$, définie par :

$${}^{tr}m(A) = \begin{cases} \frac{m(A)}{1 - m(\Omega)} & \forall A \subset \Omega, \\ 0 & \text{sinon.} \end{cases} \quad (4.42)$$

La fonction de masse ${}^{tr}m$ est obtenue à partir de m en transférant totalement et uniformément $m(\Omega)$ à tous les éléments focaux de m .

Remarque 4.8. *La matrice associée à l'opération de renforcement est donnée par :*

$$\begin{pmatrix} 1 - \alpha & 0 & \dots & 0 & 0 \\ 0 & 1 - \alpha & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & 1 - \alpha & 0 \\ \alpha & \alpha & \dots & \alpha & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{1 - \alpha} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{1 - \alpha} & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & \frac{1}{1 - \alpha} & 0 \\ \frac{-\alpha}{1 - \alpha} & \frac{-\alpha}{1 - \alpha} & \dots & \frac{-\alpha}{1 - \alpha} & 1 \end{pmatrix} \quad (4.43)$$

Cette matrice n'est pas stochastique, d'où les conditions imposées sur la fonction de masse avant la transformation, c'est-à-dire sur α dans notre cas, afin de conserver une fonction de masse après l'opération de renforcement. En particulier, on retrouve le fait que les fonctions de masse ${}^{tr}m$ et m ne sont pas distinctes.

4.7.2 Schéma étendu de Zhu et Basir [127]

Dans [127], Zhu et Basir reconnaissent cette nécessité d'étendre le processus d'affaiblissement afin de pouvoir renforcer ou « contredire » une fonction de croyance en plus de l'affaiblir.

1. Un processus pouvant affaiblir ou renforcer une source est obtenu en reprenant l'équation 4.10 de l'affaiblissement classique, et en autorisant le taux d'affaiblissement α à se tenir dans l'intervalle $[\frac{-m(\Omega)}{1 - m(\Omega)}, 1]$. Il suffit de remarquer que l'équation de l'affaiblissement classique avec $\alpha \in [\frac{-m(\Omega)}{1 - m(\Omega)}, 0]$ est une réécriture

de l'équation de renforcement de Denœux et Smets (4.41), avec une reparamétrisation $\alpha = \frac{-\alpha'}{1-\alpha'}$ où $\alpha' \in [0, m(\Omega)]$. En effet :

$${}^\alpha m = \left(1 - \frac{-\alpha'}{1-\alpha'}\right)m + \frac{-\alpha'}{1-\alpha'}m_\Omega = \frac{m - \alpha'm_\Omega}{1 - \alpha'}. \quad (4.44)$$

2. Un processus de contradiction d'une information m non vide est obtenu par l'opération suivante :

$$\begin{cases} {}^\alpha m(\bar{A}) &= (\alpha - 1)m(A) & \text{si } A \subset \Omega, \\ {}^\alpha m(\Omega) &= (\alpha - 1)m(\Omega) + 2 - \alpha & \text{sinon,} \end{cases} \quad (4.45)$$

où $\alpha \in [1, 1 + \frac{1}{1-m(\Omega)}]$. Si $\alpha = 1$, ${}^\alpha m = m_\Omega$. Si $\alpha = 1 + \frac{1}{1-m(\Omega)}$, ${}^\alpha m = \overline{tr}m$, où $\overline{tr}m$ désigne la négation de m [35], définie par $\overline{tr}m(A) = m(\bar{A})$, $\forall A \subseteq \Omega$. Autrement dit, si $\alpha = 1 + \frac{1}{1-m(\Omega)}$, chaque masse m fournie par la source est transférée à son complémentaire ; la masse est alors totalement renforcée.

Ce schéma a été appliqué avec succès en imagerie médicale [127]. Cependant il souffre d'un manque de justification formelle. En effet, le nombre $(1 - \alpha)$ ne peut plus être interprété comme un degré de croyance puisqu'il peut être supérieur à 1 ou inférieure à 0. Dans le paragraphe suivant, une nouvelle interprétation est donnée à l'équation (4.45), à l'équation (4.10) où $\alpha \in [\frac{-m(\Omega)}{1-m(\Omega)}, 1]$, ainsi qu'à l'équation du renforcement de Denœux et Smets (4.41), en même temps qu'est introduit un mécanisme général de correction de fonction de croyance.

4.7.3 Expression générale d'un mécanisme de correction

Dans l'opération d'affaiblissement, l'agent considère que la source peut se trouver dans deux états : fiable ou non fiable. L'interprétation de ces états de fiabilité est donnée par une transformation de la fonction de masse fournie sachant l'état dans lequel se trouve la source : si la source est fiable, l'agent accepte l'information, sinon il la remplace par la fonction de masse vide. Dans ce paragraphe, nous généralisons ces hypothèses, en considérant que la source peut se trouver dans N états R_i , $i \in \llbracket 1, N \rrbracket$. L'interprétation de ces états est donnée par une transformation m_i de m . Sachant que la source est dans l'état R_i : $m_{Ag}^\Omega[R_i] = m_i \forall i \in \llbracket 1, N \rrbracket$. Nous noterons \mathcal{R} l'espace composé de ces états dans lesquels peut se trouver la source, et supposons $m_{Ag}^\mathcal{R}(\{R_i\}) = \gamma_i$, $\forall i \in \llbracket 1, N \rrbracket$, avec $\sum_{i=1}^N \gamma_i = 1$.

Ainsi, la connaissance de l'agent Ag , composée de l'information m_S^Ω fournie par une source, et de la métaconnaissance $m_{Ag}^\mathcal{R}$ sur l'état dans lequel se trouve la source, peut se traduire par la formule suivante :

$$m_{Ag}^\Omega[m_S^\Omega, m_{Ag}^\mathcal{R}] = \left(\bigodot_{i=1}^N m_{Ag}^\Omega[R_i] \uparrow^{\Omega \times \mathcal{R}} \bigodot m_{Ag}^\mathcal{R} \uparrow^{\Omega \times \mathcal{R}} \right) \downarrow^\Omega. \quad (4.46)$$

Proposition 4.10. *La fonction de masse m_{Ag}^Ω , résultante de l'équation (4.46) dépend seulement des m_i et des γ_i , $i \in \llbracket 1, N \rrbracket$. Le résultat est alors noté γm , où γ désigne le vecteur composé des γ_i , et vérifie :*

$$m_{Ag}^\Omega = \gamma m = \sum_{i=1}^N \gamma_i m_i. \quad (4.47)$$

Preuve : Voir annexe A, paragraphe A.4. \square

À partir de l'équation (4.47), les équations des mécanismes de correction précédemment introduits peuvent être retrouvées.

1. L'opération d'affaiblissement classique correspond au cas où la source peut se trouver dans deux états R_1 et R_2 , tels que $m_1 = m_\Omega$ et $m_2 = m$.
2. L'opération de renforcement de Denœux et Smets, est retrouvée en considérant que la source peut se trouver dans deux états, tels que $m_1 = m$ et $m_2 = {}^{tr}m$, c'est-à-dire un état où l'information fournie par la source est acceptée, et un état où l'information fournie par la source est totalement renforcée. Avec ces hypothèses, $\gamma m = \gamma_1 m + \gamma_2 {}^{tr}m$, qui est une reparamétrisation de l'équation de renforcement (4.41) avec $\gamma_1 = \frac{m(\Omega) - \alpha}{1 - \alpha m(\Omega)}$.
3. Le schéma de Zhu et Basir, relatif au renforcement et à l'affaiblissement d'une source, équation (4.10) avec $\alpha \in [\frac{-m(\Omega)}{1 - m(\Omega)}, 1]$, est aussi une reparamétrisation de l'équation (4.47) dans le cas particulier de deux états, tels que $m_1 = m_\Omega$ et $m_2 = {}^{tr}m$, c'est-à-dire un état où la source est non fiable, et un état où la source est trop prudente et l'agent désire la renforcer. Le changement de variable est $\gamma_1 = \alpha(1 - m(\Omega)) + m(\Omega)$.
4. Enfin, le schéma de Zhu et Basir, permettant de contredire une information (équation 4.45), est encore une reparamétrisation de l'équation (4.47), en posant $\gamma_1 = (\alpha - 1)(1 - m(\Omega))$, dans un cas particulier où la source peut se trouver dans deux états, tels que $m_1 = \overline{{}^{tr}m}$ et $m_2 = m_\Omega$, c'est-à-dire un état où l'agent pense que la source ment : le contraire est vrai, et un état où l'information fournie par la source est simplement rejetée.

Les hypothèses conduisant à chacun de ces mécanismes sont récapitulées sur la figure 4.8.

Remarque 4.9. *Le schéma d'affaiblissement et de renforcement de Zhu et Basir (équation 4.10 avec $\alpha \in [\frac{-m(\Omega)}{1 - m(\Omega)}, 1]$) est donc un affaiblissement classique de ${}^{tr}m$ reparamétré, et le schéma d'opposition de Zhu et Basir (équation 4.45) est un affaiblissement classique de ${}^{tr}m$ reparamétré.*

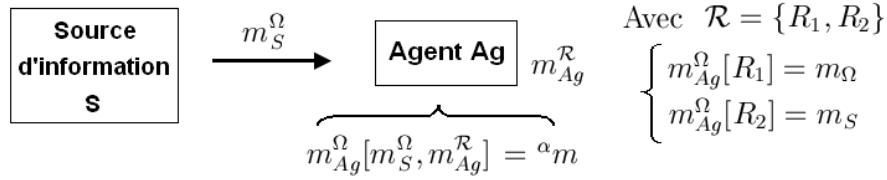
Remarque 4.10. *Avec les hypothèses conduisant à un renforcement, dans le cas où la source est trop prudente, il est possible de choisir un renforcement plus informatif de m que ${}^{tr}m$, par exemple, la distribution de probabilité pignistique ${}^{bet}m$ définie par :*

$${}^{bet}m(\{\omega\}) = \sum_{\{A \subseteq \Omega, \omega \in A\}} \frac{m(A)}{(1 - m(\emptyset))|A|}, \quad \forall \omega \in \Omega. \quad (4.48)$$

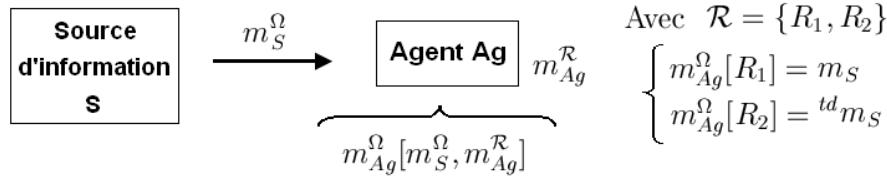
Ainsi, un autre processus de renforcement est défini par : $\gamma m = \gamma_1 m + \gamma_2 {}^{bet}m$. L'opération de renforcement de Denœux et Smets est un processus de renforcement particulier.

Remarque 4.11. *En posant $m_{Ag}^{\mathcal{R}}$ de la forme suivante :*

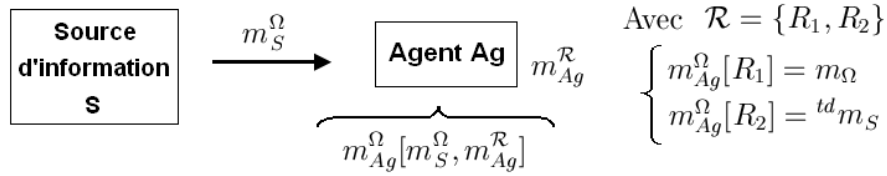
$$\begin{cases} m_{Ag}^{\mathcal{R}}(\{R_i\}) &= \gamma_i & \forall i \in \llbracket 1, N \rrbracket, \\ m_{Ag}^{\mathcal{R}}(\mathcal{R}) &= 1 - \sum_{i=1}^N \gamma_i, \end{cases} \quad (4.49)$$



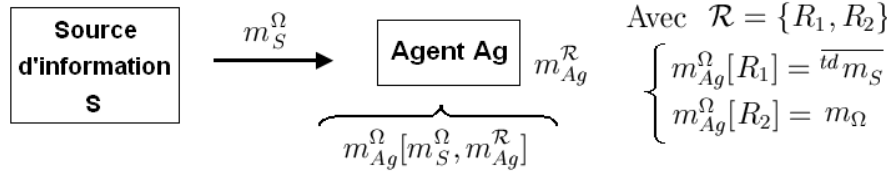
(a) Modèle de correction conduisant à l'affaiblissement classique



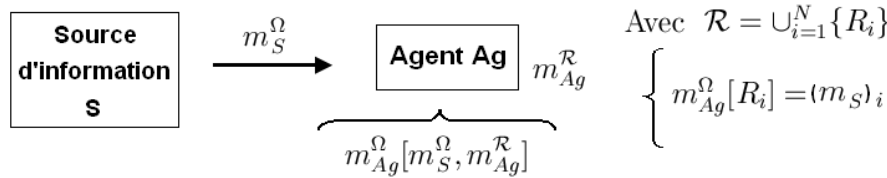
(b) Modèle de correction conduisant au renforcement



(c) Modèle de correction conduisant au schéma d'affaiblissement et de renforcement de Zhu et Basir



(d) Modèle de correction conduisant au schéma d'opposition de Zhu et Basir



(e) Définition du modèle de correction générale

Figure 4.8 – Illustrations des modèles conduisant aux mécanismes de correction introduit au paragraphe 4.7.

avec $\sum_{i=1}^N \gamma_i \leq 1$, un degré de liberté supplémentaire peut être gagné :

$$\gamma_m = \sum_{i=1}^N \gamma_i m_i + (1 - \sum_{i=1}^N \gamma_i) m_\Omega . \quad (4.50)$$

4.8 Conclusion

Dans ce chapitre, de nouveaux mécanismes de correction de fonction de croyance ont été introduits.

Dans un premier temps, nous avons travaillé sur l'expression de la croyance tenue par un agent au regard de la fiabilité d'une source d'information : celle-ci peut dépendre du contexte, de la valeur prise par la vérité. En prenant en compte ces métaconnaissances conditionnelles, l'affaiblissement contextuel a été développé. Cette nouvelle opération est paramétrée par un vecteur de taux d'affaiblissement dont les composantes correspondent à la fiabilité de la source dans chaque contexte. Ces taux peuvent être choisis par un utilisateur, ou appris à partir d'un ensemble d'apprentissage. L'affaiblissement classique est retrouvé comme cas particulier de cette opération, lorsqu'un seul taux d'affaiblissement est connu dans un contexte où la valeur de vérité se trouve bien dans Ω .

Nous avons ensuite étendu les états de fiabilité dans lesquels peut se trouver une source d'information, et introduit un mécanisme général de correction. Ce mécanisme offre une nouvelle interprétation au processus de renforcement de Denœux et Smets ainsi qu'aux schémas de Zhu et Basir, et permet de développer des stratégies de correction de fonction de croyance plus fines.

Il serait possible de contextualiser ce mécanisme de correction général comme nous l'avons fait pour l'affaiblissement classique, de même que d'essayer d'apprendre automatiquement les coefficients γ_i en fonction d'un ensemble d'apprentissage. Cela entre dans le cadre de nos futurs travaux.

Dans le prochain chapitre, seront employés l'affaiblissement contextuel sur un grossissement particulier de Ω , ainsi qu'un mécanisme de correction permettant d'affaiblir ou de renforcer une fonction de croyance.

Améliorations du modèle

Dans ce chapitre, des travaux d'amélioration du modèle de base, présenté au chapitre 3, sont exposés. Ces travaux portent principalement sur l'utilisation de l'affaiblissement contextuel, sur la prise en compte de scores de confiance, et sur l'étude de la dépendance des LAP 1 et 2 d'origine Solystic. Au cours de ce chapitre, différentes perspectives à court terme sont présentées. Celles-ci illustrent notamment la modularité des modèles déjà développés.

5.1 Utilisation de l'affaiblissement contextuel

Une première voie d'amélioration consiste à prendre en compte ce qui est potentiellement reconnaissable ou non, par chacun des LAP à combiner. Dans ce paragraphe, nous montrons comment l'information fournie par un LAP ne travaillant pas sur l'espace complet des décisions peut être pris en compte par un affaiblissement contextuel. Cet affaiblissement est réalisé sur la fonction de masse ayant été initialisée par l'affectation postale exposée au paragraphe 3.2.3. Il est basé sur un grossissement de Ω composé de deux éléments :

- un premier regroupant les éléments que le LAP ne sait pas reconnaître ;
- un second composé des éléments que le LAP sait reconnaître.

Le vecteur d'affaiblissement est alors donné par le couple $(1, 0)$:

- le taux d'affaiblissement est égal à 1 pour les éléments non reconnus par le LAP : sachant que la vérité est un élément que le LAP ne sait pas reconnaître, le LAP ne la décidera pas ;
- le taux d'affaiblissement est égal à 0 pour les éléments reconnus par le LAP : l'affectation postale est considérée comme adéquate pour la détection de ces éléments.

Dans notre application, les LAP 1 et 2 peuvent tout reconnaître, seul le LAP 3 ne reconnaît pas les distributions spéciales, c'est-à-dire les boîtes postales (cf paragraphe 1.1.1). Ainsi, le grossissement Θ de Ω , illustré sur la figure 5.1, est employé pour affaiblir la fonction de masse issue de ce LAP. Ce grossissement comprend deux éléments : un premier regroupant toutes les distributions spéciales et un second regroupant tous les autres éléments de Ω .

La combinaison conjonctive des fonctions de masse issues des LAP 1 et 2 avec cet affaiblissement contextuel de la fonction de masse issue du LAP 3 fournit, pour les mêmes réglages de coûts R_1 à R_{15} , les mêmes points de fonctionnement au niveau distribution que la combinaison de base exposée au chapitre 3. Au niveau

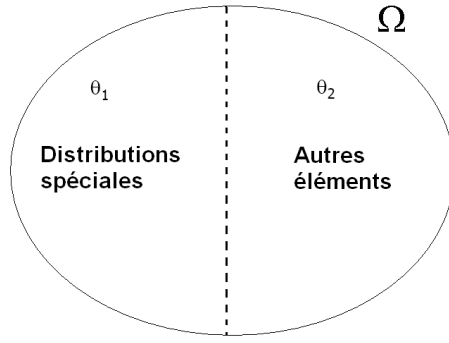


Figure 5.1 – Éléments du grossissement employé pour réaliser un affaiblissement Θ -contextuel de la fonction de masse issue du LAP 3.

acheminement, quelques bonnes reconnaissances supplémentaires sont obtenues pour une raison qui sera détaillée un peu plus loin.

En examinant les résultats, on constate que pour chaque distribution spéciale identifiée par les LAP Solystic aboutit à la même *situation*, c'est-à-dire au même triplet similaire de décisions (cf remarque 3.3). Les LAP Solystic 1 et 2 décident la même distribution spéciale « $A_i B_j$ », et le LAP 3 fournit la décision partielle acheminement : « $A_i Rej$ ». Autrement dit, le LAP 3 ne reconnaît pas la distribution mais ne se trompe pas d'acheminement. Pour chaque triplet de décisions de ces catégories, en ce qui concerne les envois manuscrits, les fonctions de masse suivantes sont obtenues :

$$\begin{cases} m_1(A_i) & = 0.019737 \\ m_1(\{A_i B_j\}) & = 0.980263 \end{cases} \quad \begin{cases} m_2(A_i) & = 0.038961 \\ m_2(\{A_i B_j\}) & = 0.961039 \end{cases}$$

$$\begin{cases} m_3(\Omega) & = 0.009512 \\ m_3(inv) & = 0.000607 \\ m_3(A_i) & = 0.989881, \end{cases}$$

où m_i désigne la masse issue du LAP i , pour tout $i \in \{1, 2, 3\}$. À chaque fois que les LAP 1 et 2 fournissent une décision de catégorie « distribution spéciale », ce même triplet de décisions est retrouvé. Cette situation concerne environ 150 images manuscrites sur les 6000 images manuscrites distribuables du lot d'apprentissage, composé de 28000 lettres. Nous pouvons remarquer le caractère engagé des masses traduisant la bonne performance des décisions de cette catégorie issues des LAP Solystic. Dans cette même configuration de décisions dans le cas des envois dactylographiés, les fonctions de masses sont encore plus engagées.

Sans l'affaiblissement contextuel, le calcul de $m = m_1 \odot m_2 \odot m_3$ donne :

$$\begin{aligned} m(A_i) & = 0.000769 \\ m(\{A_i B_j\}) & = 0.998624 \\ m(\emptyset) & = 0.000607 . \end{aligned} \tag{5.1}$$

Avec $\gamma_1 = \{A_i B_j\}$, $\gamma_2 = \{A_i inv\}$, et $\gamma_3 = A_i \setminus \{A_i B_j, A_i inv\}$, la probabilité pignistique sur Γ est donnée par :

$$\begin{aligned} BetP^\Gamma(\{\gamma_1\}) &= \mathcal{K}(\frac{0.000769}{3} + 0.998624) = 0.99948702 \\ BetP^\Gamma(\{\gamma_2\}) &= \mathcal{K}\frac{0.000769}{3} = 0.00025649 \\ BetP^\Gamma(\{\gamma_3\}) &= \mathcal{K}\frac{0.000769}{3} = 0.00025649 \end{aligned} \tag{5.2}$$

avec $\mathcal{K} = \frac{1}{1-m(\emptyset)} = 1.000607$.

En considérant les coûts choisis pour l'obtention des points de fonctionnement illustrés sur les figures 3.8 à 3.11, obtenus par les réglages R_1 à R_{15} , la probabilité pignistique portant sur la distribution spéciale est déjà suffisamment importante pour que le risque pignistique associé à cette décision soit toujours le plus faible, et ainsi, pour que cette décision soit toujours fournie par la combinaison. L'information contextuelle améliore les performances de la combinaison au niveau distribution dans le cas de réglages plus sévères sur les erreurs, réglages n'ayant pas été sélectionnés au chapitre 3. La figure 5.2 illustre à nouveau les performances de la combinaison en distribution en ajoutant deux nouveaux points, obtenus avec un même réglage R_0 , l'un sans l'affaiblissement contextuel du LAP 3, le deuxième avec affaiblissement contextuel du LAP 3. Avec cet affaiblissement contextuel, le même point de fonctionnement est obtenu par le réglage R_0 et par le réglage R_1 associé au point de fonctionnement nommé $CMCT_1$, qui demeure à la même position avec ou sans l'affaiblissement contextuel. Le gain de bonnes reconnaissances obtenu entre le réglage R_0 sans affaiblissement contextuel, et le réglage R_0 avec affaiblissement contextuel, correspond aux 150 images de la situation évoquée dans ce paragraphe.

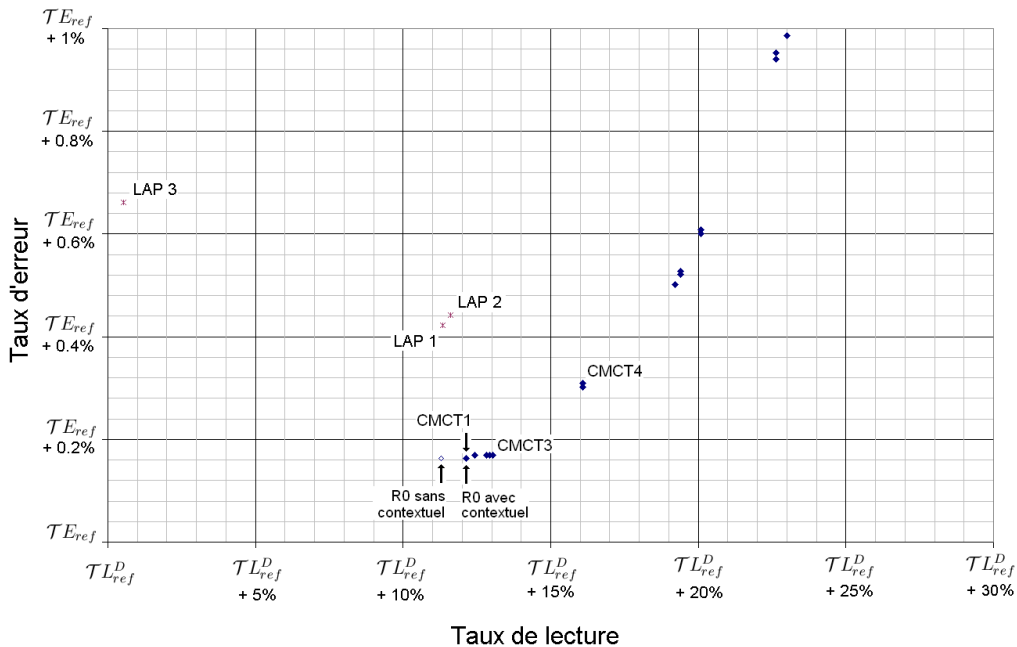


Figure 5.2 – Illustration de l'amélioration au **niveau distribution** d'un point de fonctionnement obtenu grâce à un affaiblissement contextuel basé sur le grossissement illustré sur la figure 5.1.

Néanmoins il est possible de tirer avantage de cette information contextuelle au niveau acheminement. En effet, avec l'affaiblissement contextuel, le calcul de $m = m_1 \odot m_2 \odot^{(1,0)} m_3$ donne :

$$\begin{aligned} m(A_i) &= 0.000769 \\ m(\{A_i B_j\}) &= 0.999231 . \end{aligned} \quad (5.3)$$

La masse sur le vide est nulle. Le conflit présent dans l'équation (5.1) n'existe plus, il a été expliqué par l'information de non reconnaissance des distributions spéciales de la part du LAP 3. Ce résultat est plus engagé en faveur de la distribution spéciale. Avec cet affaiblissement, la probabilité pignistique est donnée par :

$$\begin{aligned} BetP^\Gamma(\{\gamma_1\}) &= \left(\frac{0.000769}{3} + 0.999231\right) = 0.99948734 \\ BetP^\Gamma(\{\gamma_2\}) &= \frac{0.000769}{3} = 0.00025633 \\ BetP^\Gamma(\{\gamma_3\}) &= \frac{0.000769}{3} = 0.00025633 \end{aligned} \quad (5.4)$$

avec $\gamma_1 = \{A_i B_j\}$, $\gamma_2 = \{A_i inv\}$, et $\gamma_3 = A_i \setminus \{A_i B_j, A_i inv\}$. Ainsi, avec un coût de rejet distribution plus faible, le risque pignistique de la décision associée à cette distribution spéciale sera toujours assez faible pour que cette décision soit fournie. Un coût plus faible de rejet en distribution, à coût d'erreur constant, diminue le risque pignistique attaché aux décisions de niveau acheminement. Celles-ci pourront alors être prises alors qu'elles étaient précédemment rejetées. Ainsi, avec le réglage R_0 et l'affaiblissement contextuel, nous obtenons le même point de fonctionnement $CMCT_1$ en distribution, et un nouveau point de fonctionnement en acheminement améliorant ce premier point de fonctionnement $CMCT_1$, comme illustré sur la figure 5.3. Sur ce lot de courrier, ce gain s'élève à 24 bonnes reconnaissances. En observant que le nombre de bonnes reconnaissances déjà fournies par le réglage $CMCT_1$ est de l'ordre de 15400 en acheminement dactylographié (figure 3.10(a)), et de l'ordre de 9900 en acheminement manuscrit (figure 3.10(a)), cet apport demeure marginal. Néanmoins, cet affaiblissement semble toujours bénéfique. Au pire, les points de fonctionnement demeurent inchangés. Cette information apparaît déjà dans la conversion par les matrices de confusion.

Dans ce paragraphe, nous avons montré comment une information de non reconnaissance d'une partie de l'univers de la part d'un des LAP pouvait être prise en compte, et comment cette information a permis d'améliorer les performances des points de fonctionnement prudents en termes de taux d'erreur. Néanmoins, notre objectif principal demeure l'obtention d'un taux de lecture maximal associé à un taux d'erreur inférieur à l'un des taux d'erreur des LAP Solystic. Le paragraphe suivant expose une des pistes que nous avons suivie pour atteindre cet objectif.

5.2 Prise en compte de scores de confiance

Nous avons eu la possibilité d'utiliser des scores de confiance pour les LAP 1 et 2 d'origine Solystic, pour ce qui concerne les envois manuscrits. Chaque décision issue de ces LAP est accompagnée d'un score compris entre 0 et 2000, calculé à partir des différents algorithmes et chaînes de décision employés. La sémantique de ces scores

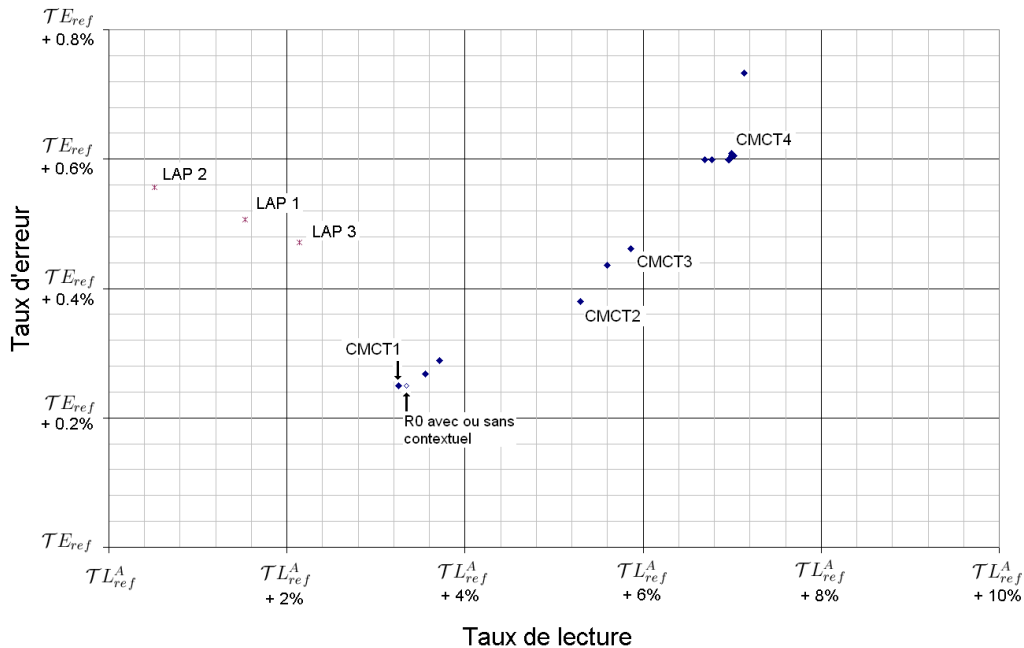


Figure 5.3 – Performances au **niveau acheminement**. Illustration d’un réglage R_0 plus intéressant que le réglage R_1 associé au point $CMCT_1$, et qui, grâce à l’affaiblissement contextuel, obtient les mêmes performances au niveau distribution que le point $CMCT_1$ (figure 5.2).

est la suivante : « plus le score est élevé, plus la confiance du LAP en la solution est élevée ».

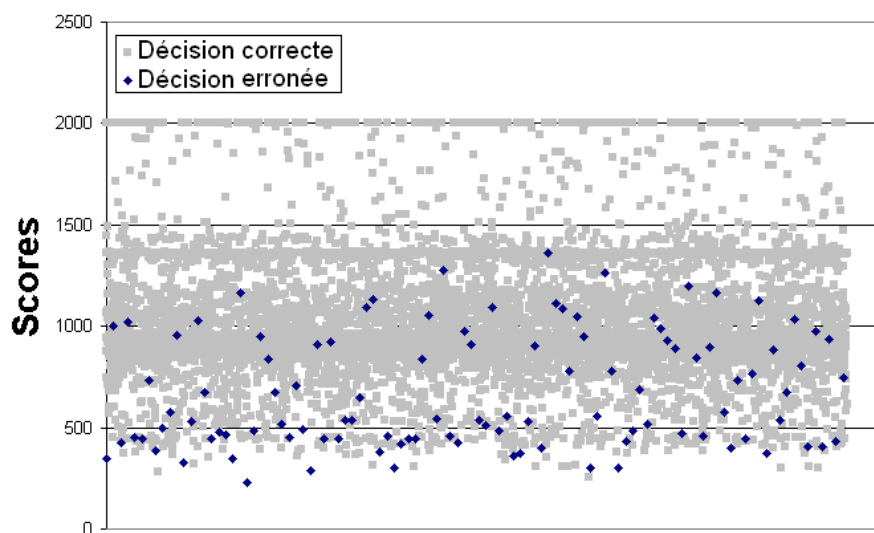
Afin d’illustrer l’intérêt de l’information apportée par ces scores, les décisions fournies par les LAP Solystic sur l’ensemble d’apprentissage, correctes ou erronées au niveau acheminement, sont représentées sur les figures 5.4(a) et 5.4(b). L’axe des ordonnées indique le score associé à une décision, l’axe des abscisses n’ayant pour rôle que de faciliter la représentation.

Sur ce lot de courrier, la sémantique des scores se vérifie pour les deux LAP. Plus le score est élevé, plus la proportion de décisions correctes est importante. Des informations plus fines peuvent aussi être observées. Par exemple, toutes les décisions ayant un score supérieur à 1500 sont correctes.

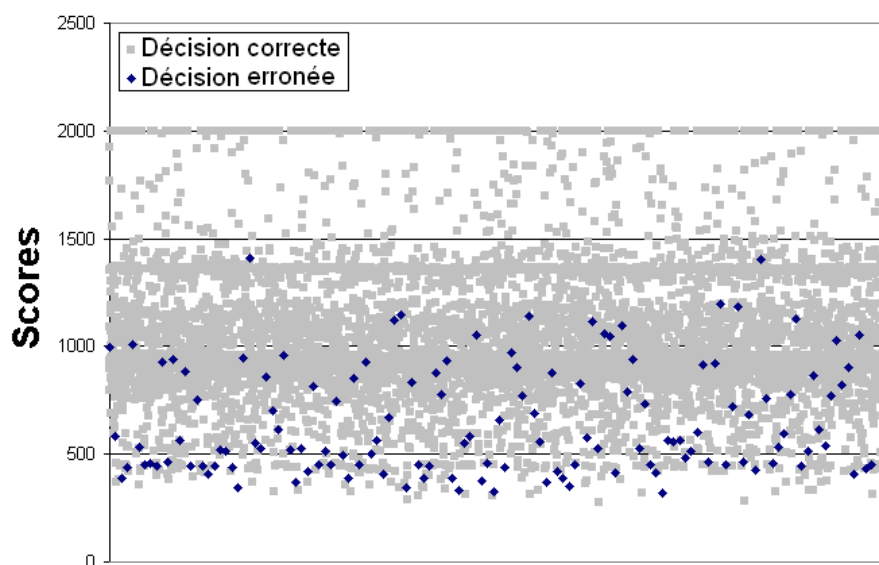
Nous proposons d’interpréter ces scores comme des paramètres de correction de l’affectation postale introduite au chapitre 3 (figure 5.5). Comme nous l’avons observé dans la remarque 3.3, la conversion des décisions d’une même catégorie en une fonction de masse est réalisée identiquement par l’affectation postale employée. Chaque triplet similaire de décisions est traité de façon identique par le modèle de base. Un des intérêts de ces scores consiste à apporter une information intrinsèque à chaque image, permettant ainsi d’individualiser chacune de ces situations.

Deux approches ont été testées, l’une avec un affaiblissement classique, l’autre avec un mécanisme général de correction.

La première approche consiste à apprendre, automatiquement et individuellement (cf paragraphe 4.6.1), un taux d’affaiblissement par une variante de l’approche



(a) Décisions fournies par le LAP 1.



(b) Décisions fournies par le LAP 2.

Figure 5.4 – Scores et vérité au **niveau acheminement** des décisions manuscrites, différentes d'un rejet complet, fournies par les LAP 1 et 2 sur le lot de courrier d'apprentissage. Un losange foncé correspond à une décision erronée au niveau acheminement. Un carré clair est associé à une décision correcte au niveau acheminement.

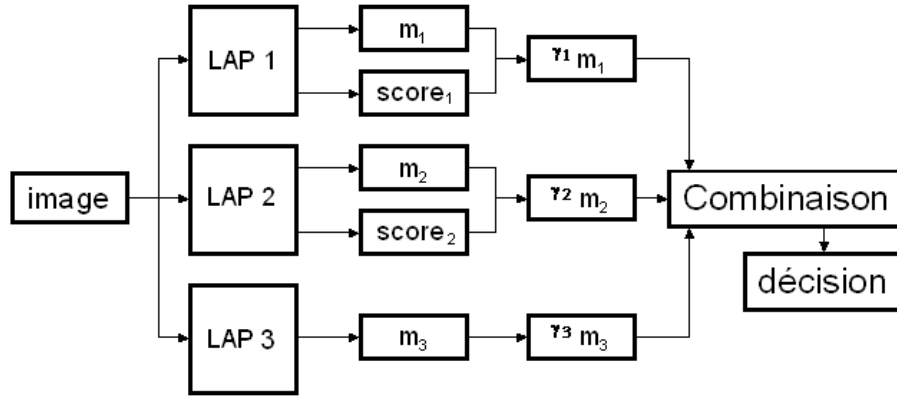


Figure 5.5 – Extension du modèle de base par la prise en compte de scores de confiance comme paramètres de correction de l’affectation postale employée.

proposée par Elouedi et al [40]. L’idée consiste à calculer, pour chaque LAP Solystic, un taux d’affaiblissement, fonction du score associé à la décision fournie par le LAP. Dans le critère à minimiser $E_{bet}(\alpha)$, équation (4.37), le taux d’affaiblissement α peut par exemple être remplacé par la fonction $\exp(-c.s_i)$, où s_i est le score attaché à la décision du LAP au regard de l’objet o_i et c est la constante à apprendre (figure 5.6(a)). Un nouveau critère à minimiser devient :

$$E_{bet}(c) = \sum_{i=1}^n \sum_{k=1}^K (\alpha_i BetP^\Gamma\{o_i\}(\omega_k) - \delta_{i,k})^2, \quad (5.5)$$

où $\alpha_i = \exp(-c.s_i)$ et $\alpha_i BetP^\Gamma\{o_i\}$ est la probabilité pignistique sur le cadre de pari Γ (cf algorithme 1, chapitre 3) attachée à la fonction de masse $m^\Omega\{o_i\}$ affaiblie par le taux d’affaiblissement α_i :

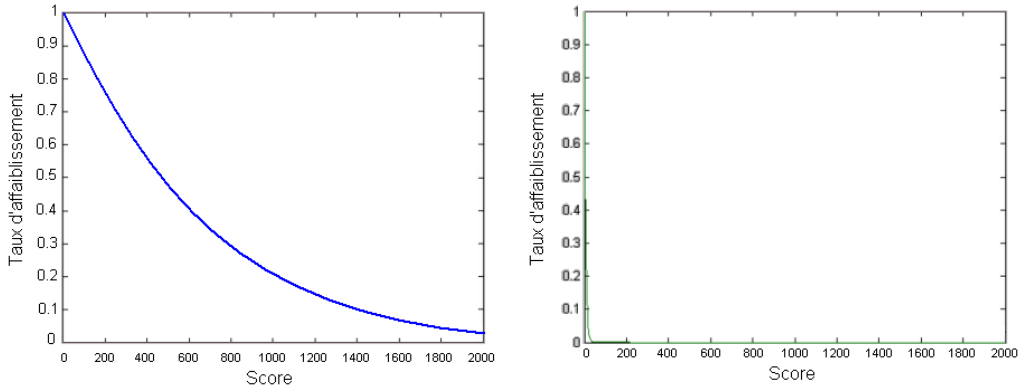
$$\alpha_i m^\Omega\{o_i\} = (1 - \exp(-c.s_i))m^\Omega\{o_i\} + \exp(-c.s_i) m_\Omega. \quad (5.6)$$

Pour les deux LAP Solystic, le résultat de la minimisation de ce critère par l’emploi d’une procédure générale d’optimisation non linéaire sous la contrainte $c \geq 0$, est illustré sur la figure 5.6(b). Ce résultat peut être interprété de la manière suivante : par le procédé de correction employé et l’intervention tolérée du score, il est préférable de ne pas modifier les fonctions de croyance.

En poursuivant cette même idée, un apprentissage collectif (cf paragraphe 4.6.2) peut aussi être réalisé. Il consiste à minimiser, sous les contraintes $c_1 \geq 0$, $c_2 \geq 0$, et $0 \leq \alpha_3 \leq 1$, le critère suivant :

$$E_{bet}(c_1, c_2, \alpha_3) = \sum_{i=1}^n \sum_{k=1}^K (BetP_{123}^\Gamma\{o_i\}(\omega_k) - \delta_{i,k})^2, \quad (5.7)$$

où $\alpha_{ij} = \exp(-c_j.s_i)$, $j \in \{1, 2\}$, et $BetP_{123}^\Gamma\{o_i\}$ est la probabilité pignistique issue de la combinaison conjonctive des trois fonctions de masses affaiblies $\alpha_{i1} m_1^\Omega\{o_i\}$, $\alpha_{i2} m_2^\Omega\{o_i\}$, et $\alpha_{i3} m_3^\Omega\{o_i\}$ issues respectivement des LAP 1, 2 et 3. Par l’emploi d’une



(a) Forme cherchée pour la fonction des scores.

(b) Fonction apprise.

Figure 5.6 – Apprentissage d’une correction des fonctions de masse issues des LAP 1 et 2 par un affaiblissement classique avec un taux d’affaiblissement fonction du score attaché à la décision du LAP.

procédure générale d’optimisation non linéaire, on obtient les mêmes résultats que ceux présentés sur la figure 5.6(b) pour les LAP Solystic, une valeur très voisine de 0 étant associée à la constante α_3 . Autrement dit, le taux d’affaiblissement associé au LAP extérieur est nul, les informations qu’il fournit ne doivent pas être affaiblies. Par ce procédé de correction, il est souhaitable de ne pas modifier l’affectation des masses. Ce résultat n’est celui qui était attendu. Toutefois, on constate qu’en attribuant manuellement différentes valeurs d’un affaiblissement, la valeur du critère d’erreur augmente. Peut-être un autre critère d’erreur pourrait-il être employé. Notons que dans notre application, les erreurs sont très minoritaires par rapport aux bonnes reconnaissances. Peut-être ceci joue-t-il aussi un rôle dans l’obtention de ce résultat. Ce travail demeure à être investiguer.

Dans un second temps, en nous appuyant sur les mécanismes généraux de correction présentés au paragraphe 4.7, nous avons choisi d’appliquer le mécanisme de correction d’équation

$$\nu m = \nu_1 m_\Omega + \nu_2 m + \nu_3^{tr} m, \quad (5.8)$$

permettant d’affaiblir ou de renforcer les fonctions de masses issues des LAP 1 et 2. Les paramètres de ce mécanisme de correction, notés « γ_i » dans le chapitre 4, sont ici notés « ν_i » afin de ne pas les confondre avec les éléments du cadre de pari Γ .

S’inspirant des répartitions des décisions correctes et erronées fournies par les LAP Solystic en fonction des scores (5.7(a)), une fonction définissant les paramètres de correction ν_i , $i \in \{1, 2, 3\}$ a été construite comme indiqué sur la figure 5.7(b). Lorsque le score de la décision est inférieur à une première borne B_1 , la décision ne peut être prise en compte et un affaiblissement total de la fonction de masse est réalisé. Au contraire, si le score est supérieur à une borne B_4 , alors cette décision doit être renforcée au maximum. Si le score se trouve entre ces bornes B_1 et B_4 , on affaiblit ou renforce en fonction d’une borne B_3 . La phase d’affaiblissement est séparée en deux affaiblissements plus ou moins sévères, contrôlés par une borne B_2 associée à une valeur d’affaiblissement P . Le calcul des scores étant effectué de

manière similaire dans les deux LAP Solystic, le même mécanisme de correction est appliqué avec les mêmes paramètres. Ce paramétrage commun se justifie aussi par l'observation des répartitions des décisions erronées en fonction des scores illustrées sur la figure 5.4, qui se révèlent très proches pour les deux LAP.

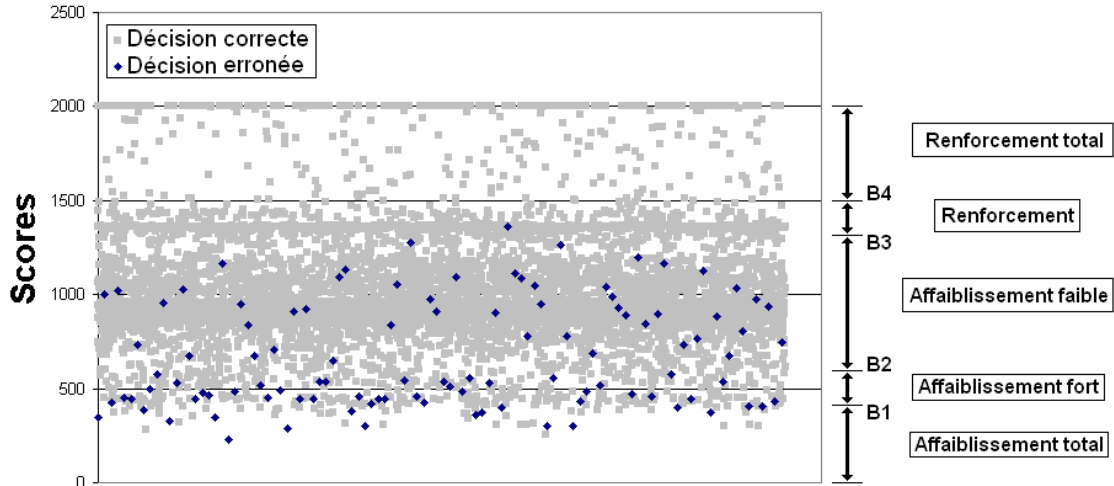
Différentes valeurs pour les paramètres B_1 , B_2 , B_3 , B_4 , et P , ont été testées manuellement. En fixant la valeur des bornes B_1 , B_2 , B_3 , B_4 , respectivement à 400, 600, 1400 et 1500, et P à 0.05, la combinaison conjonctive de la fonction de masse issue du LAP extérieur avec les fonctions de masse issues des LAP Solystic corrigées par ce mécanisme, fournit des performances probantes au niveau acheminement, comme le montre la figure 5.8(a). Ces améliorations s'observent aussi en considérant l'ensemble du courrier (figure 5.8(b)), démontrant ainsi une amélioration très encourageante du modèle de base. La proportion de courrier manuscrit est d'environ 40% sur les lots de courrier considérés.

En testant différents jeux de paramètres, nous avons observé que pour une valeur de la borne B_2 supérieure à 400, l'augmentation de la valeur de P n'était pas profitable à la combinaison. Ceci corrobore le résultat obtenu par le premier apprentissage par affaiblissement classique (figure 5.6(b)). Entre les bornes B_2 et B_3 , l'utilisation des scores par affaiblissement n'améliore pas l'affectation postale.

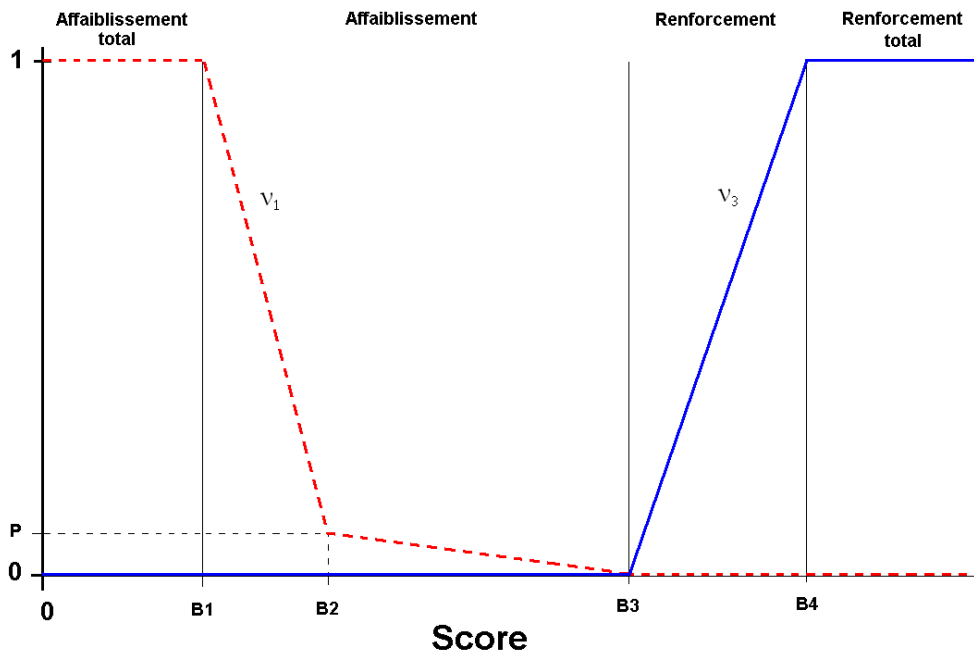
Les résultats obtenus en distribution, présentés sur la figure 5.9, sont d'apparence modeste mais demeurent très intéressants. En effet, avant l'apport des scores, le meilleur point pouvant être sélectionné était le point $CMCT_2$ ou $CMCT_3$ (cf paragraphe 3.4.3). Avec l'introduction des scores, un meilleur point peut être sélectionné : le point nommé $CMCT+$ sur les figures 5.9 et 5.8(b). Ce point $CMCT+$ est associé à un meilleur taux de lecture en acheminement et en distribution, et à un taux d'erreur inférieur au plus grand des taux d'erreur des LAP Solystic 1 et 2. Grâce aux des scores, il est ainsi possible d'obtenir de très bons compromis entre performance en acheminement et en distribution. L'ensemble de ces performances se vérifie sur l'ensemble de test, comme l'illustre la figure 5.10.

Nous pensons qu'il est possible d'améliorer encore les performances en distribution sans diminuer les acquis en acheminement. En effet, en examinant les scores fournis sur des décisions en distribution géographique présentés sur les figures 5.11(a) et 5.11(b), nous pouvons observer que les scores des décisions en distribution dans les catégories « voie non numérotée » et « pas de porte » ont des dynamiques différentes des scores sur lesquels la correction a été construite (figure 5.4). Par exemple, une décision de catégorie « pas de porte » associée à un score de 1100 semble généralement correcte, alors qu'une décision de catégorie « voie non numérotée » associée à ce même score semble beaucoup moins sûre.

Afin d'améliorer cette combinaison au niveau distribution, deux solutions sont d'ores et déjà envisageables. La première approche consiste à appliquer un coefficient multiplicateur de façon à obtenir un score entre 0 et 2000, pour chaque type de décision. Toutefois, il nous semble que nous pourrions perdre de l'information par cette normalisation. En effet, suivant le niveau auquel la décision a été fournie, le score n'a pas été construit de la même façon. Plus la décision est fine, plus grand est le nombre d'éléments ayant permis de déterminer la valeur du score. Aussi, une autre approche peut être envisagée, consistant à appliquer un mécanisme de correction fonction de la catégorie de la décision fournie. Ce n'est pas une correction

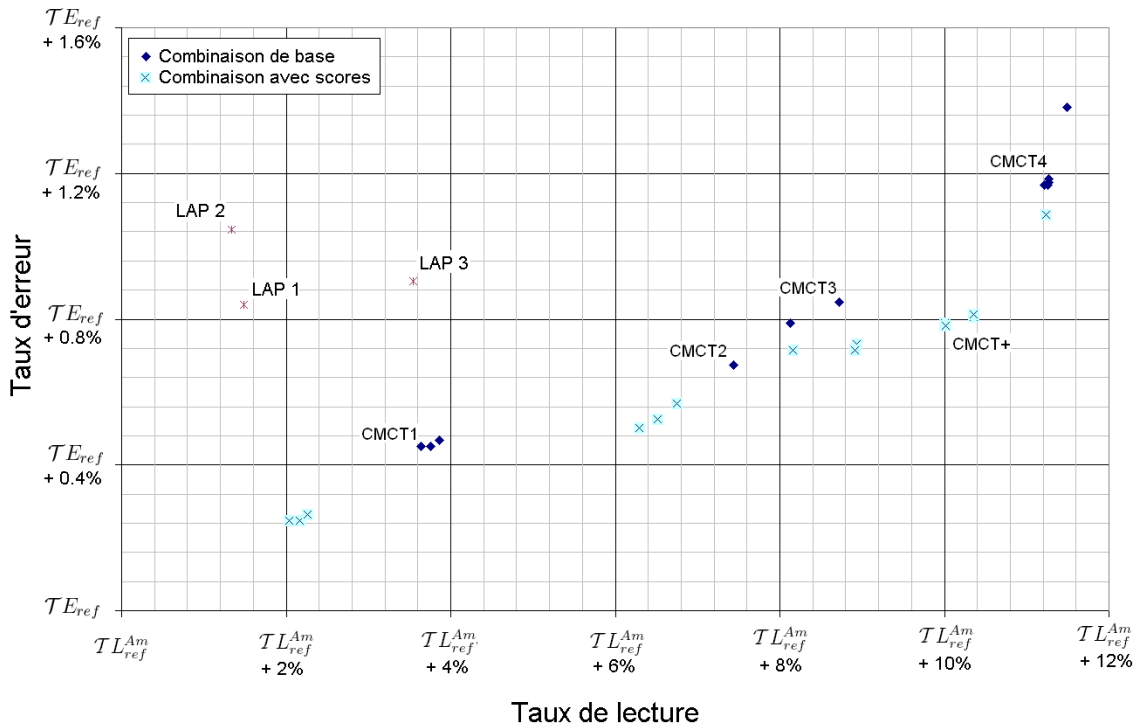


(a) Correction voulue en fonction des scores, qui sont des indicateurs pertinents des bonnes reconnaissances des LAP Solystic comme déjà illustré sur la figure 5.4.

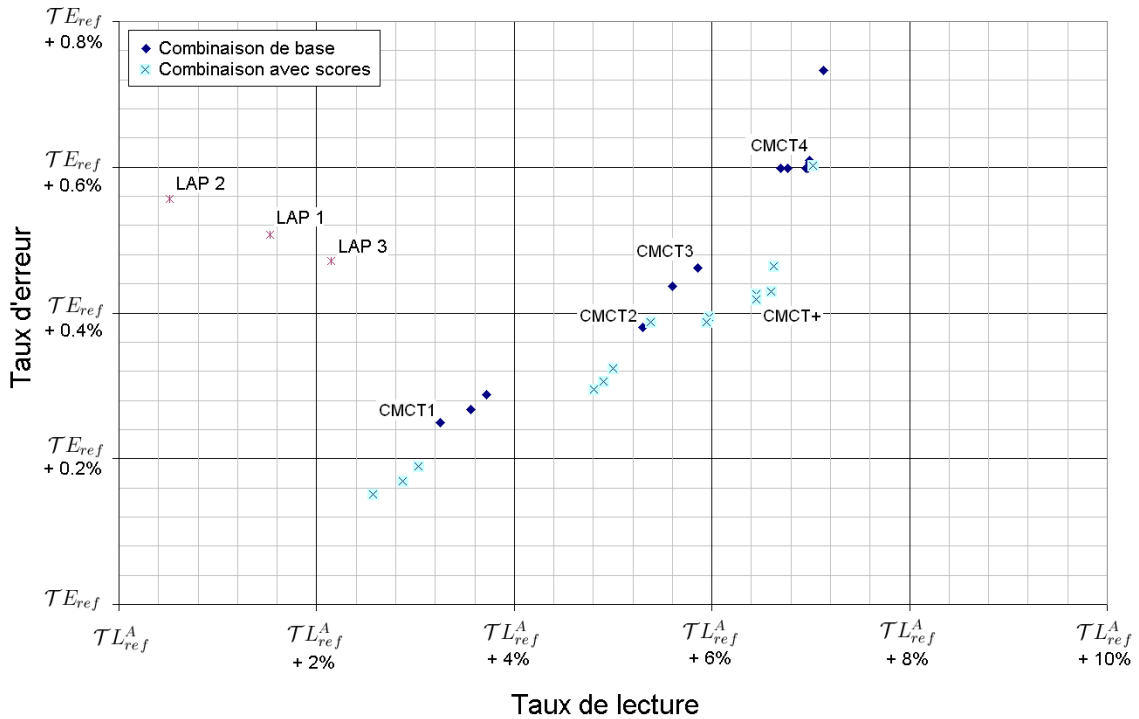


(b) Valeurs des paramètres de correction ν_1 et ν_3 du mécanisme de correction (5.8) en fonction des scores. La valeur du paramètre ν_2 est égale à $1 - \nu_1 - \nu_3$.

Figure 5.7 – Introduction d'un mécanisme de correction des fonctions de masse issues des LAP 1 et 2.



(a) Au niveau acheminement en ne considérant que les envois manuscrits. Le nombre $T_{L_{ref}}^{Am}$ indique le taux de référence en lecture au niveau acheminement en ne considérant que les envois manuscrits. Sa valeur est supérieure à 75%.



(b) Au niveau acheminement en considérant tous les envois. Rappelons que le nombre $T_{L_{ref}}^A$ indique le taux de référence en lecture au niveau acheminement.

Figure 5.8 – Performances de la combinaison de base (losange foncé) et de la combinaison avec une correction par les scores des informations issues des LAP 1 et 2 (croix sur fond clair), réalisées sur l'ensemble d'apprentissage.

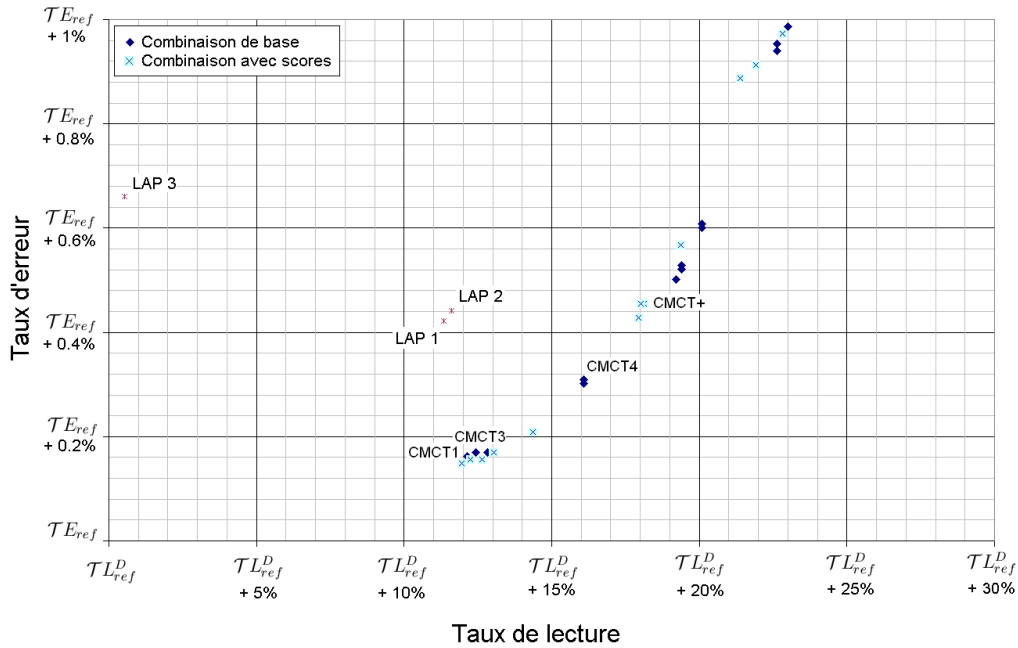
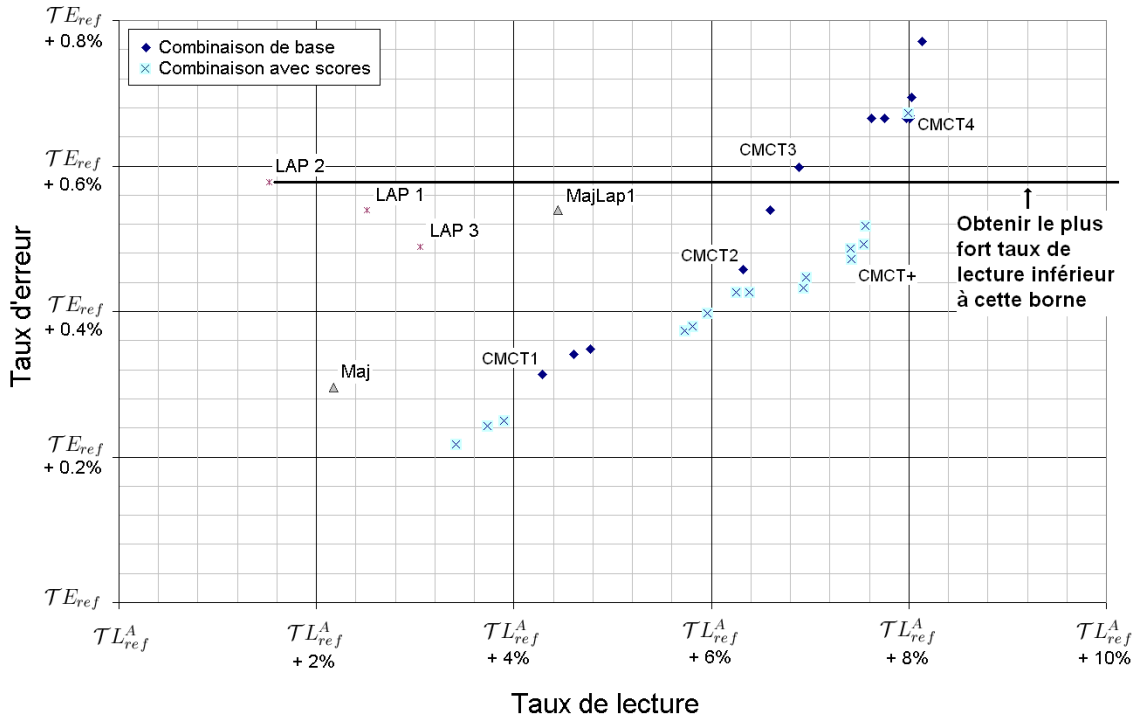


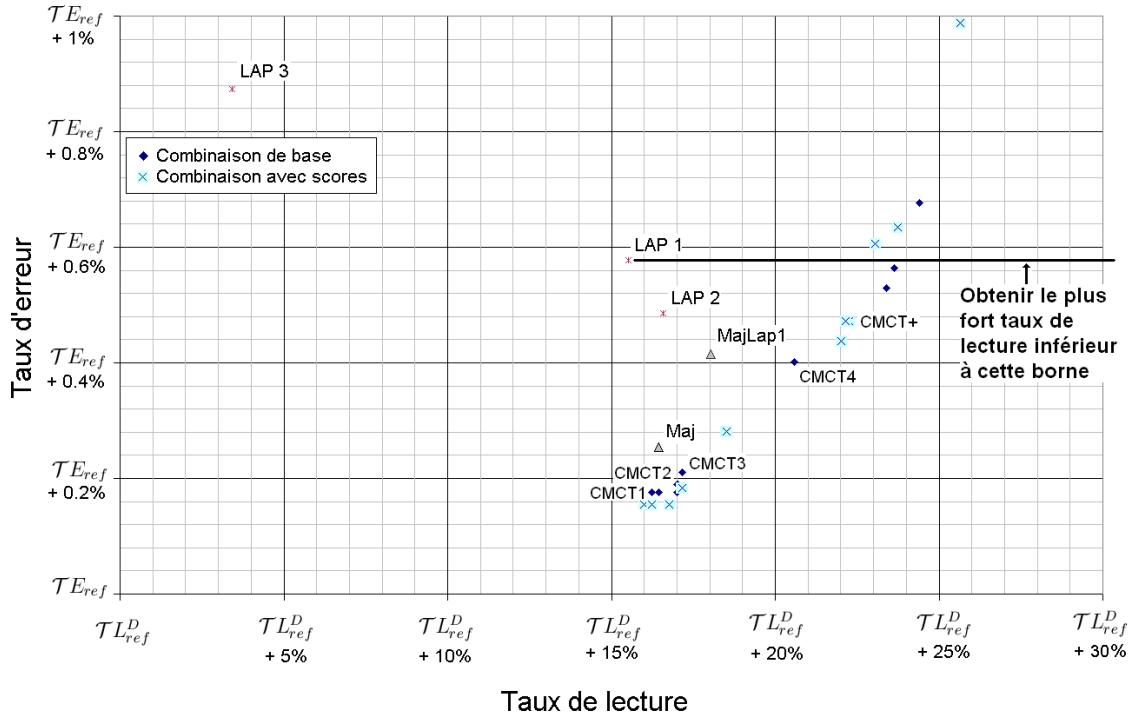
Figure 5.9 – Performances au **niveau distribution** de la combinaison de base (losange foncé) et de la combinaison avec une correction par les scores des informations issues des LAP 1 et 2 (croix sur fond clair), réalisées sur l'**ensemble d'apprentissage**.

contextuelle au sens de ce qui a été présenté au chapitre 4. Ce mécanisme de correction serait conditionnel à la catégorie de la décision fournie, il dépendrait de la décision fournie et non de la valeur de vérité de l'envoi. Un cas particulier pourrait être un mécanisme de correction hiérarchique : le vecteur de correction pourrait être composé de taux de correction conditionnels à la catégorie dans la hiérarchie de la décision fournie.

Remarque 5.1 (Retour sur le problème des sauts). *La figure 5.12(b) représentent les nombres de bonnes reconnaissances et d'erreurs en distribution manuscrite pour chaque réglage de la combinaison avec une prise en compte des scores selon le mécanisme de correction (5.8). Le résultat obtenu par la combinaison sans prise en compte des scores est rappelé sur la figure 5.12(a). Nous pouvons remarquer une atténuation notable du premier saut entre les réglages R_6 et R_7 . Comme indiqué dans la remarque 3.3, ce saut était dû à l'acceptation ou non de la décision distribution d'un des LAP Solystic alors que les deux autres LAP fournissaient une même décision de niveau acheminement en accord avec la décision de ce LAP. En fonction du score, cette décision n'est pas toujours fournie par la combinaison. En revanche, dans la situation duale où le LAP 3 fournit une décision de niveau distribution alors que les LAP Solystic fournissent une décision en accord de niveau acheminement, les scores des LAP Solystic n'apportant pas d'information sur la partie distribution, et le LAP 3 ne fournissant pas de score, le saut entre les réglages R_8 et R_9 demeure. La variation de la masse sur l'acheminement en fonction des scores des décisions des LAP Solystic ne semble pas avoir une influence suffisante. Dans le paragraphe 5.4, nous présentons de possibles éléments de solutions à ce problème.*



(a) Performances au niveau acheminement



(b) Performances au niveau distribution

Figure 5.10 – Performances de la combinaison de base (losange foncé) et de la combinaison avec une correction par les scores des informations issues des LAP 1 et 2 (croix sur fond clair), réalisées sur le lot de courrier de test.

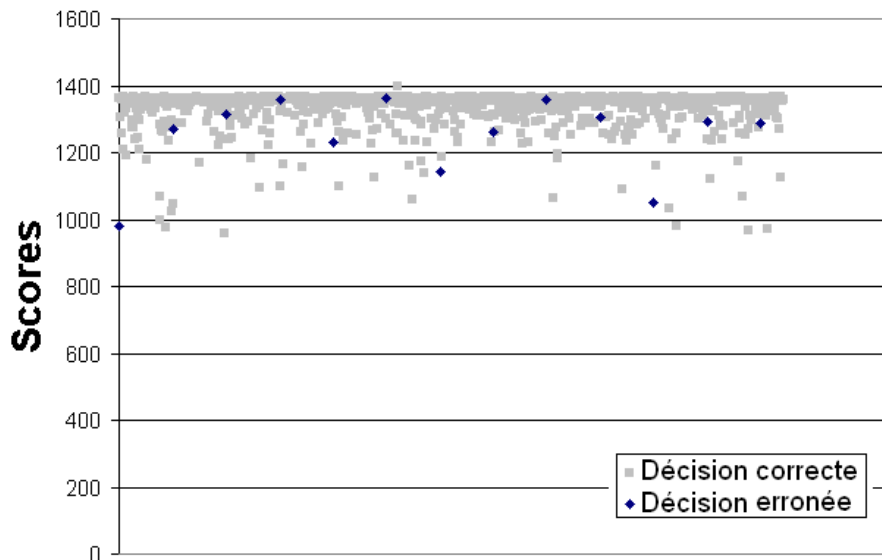
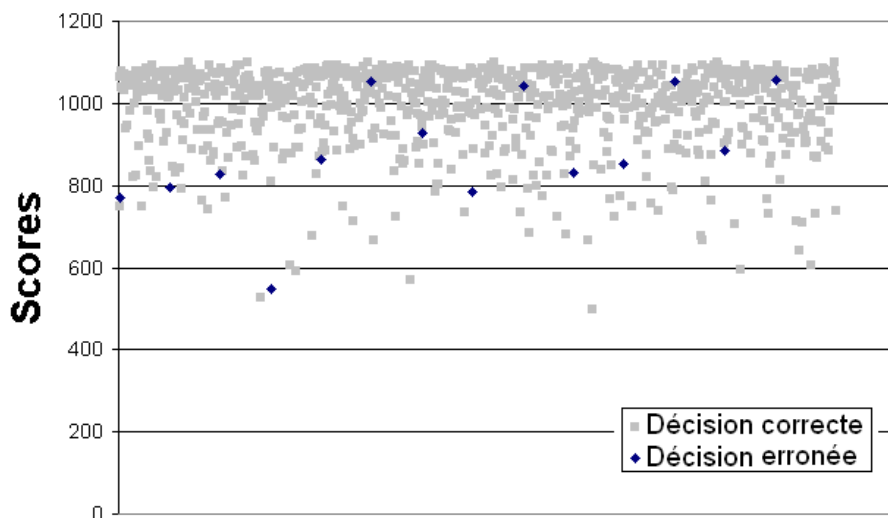
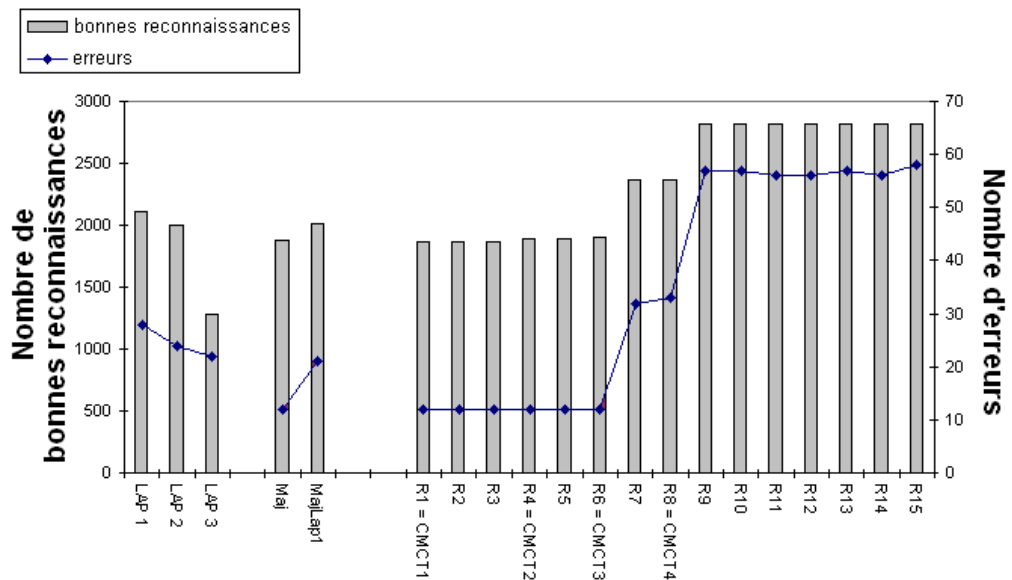
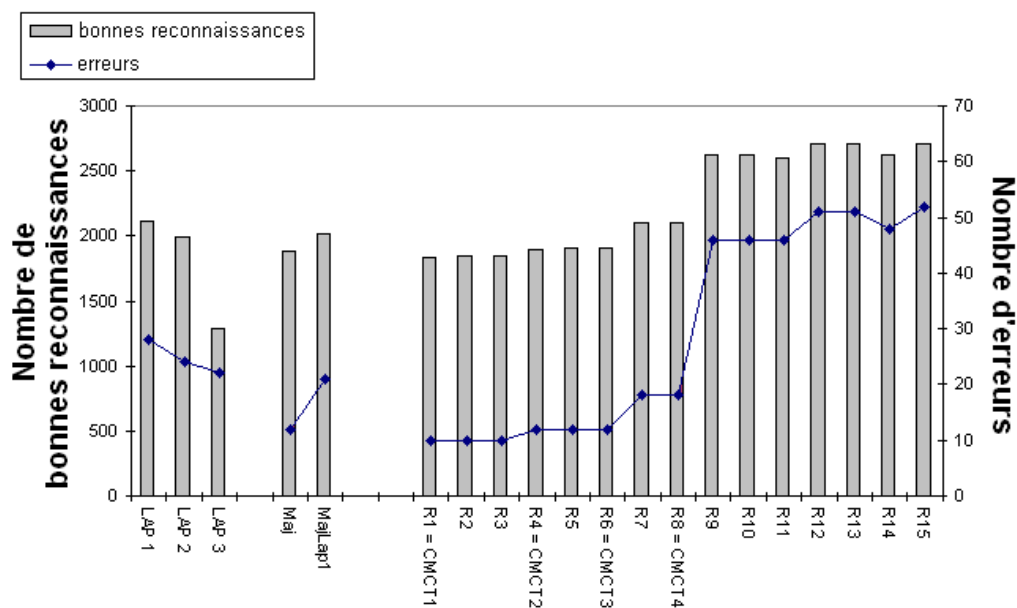
(a) Décisions de catégorie **voie non numérotée**.(b) Décisions de catégorie **pas de porte**.

Figure 5.11 – Scores et vérité au **niveau distribution** des décisions manuscrites fournies par le LAP 1 sur le lot de courrier d'apprentissage. Un losange foncé correspond à une décision erronée au niveau distribution. Un carré clair est associé à une décision correcte au niveau distribution.



(a) Sans intervention des scores (équivalent de la figure 3.11(b)).



(b) Avec prise en compte des scores selon le mécanisme (5.8).

Figure 5.12 – Nombres de bonnes reconnaissances et d’erreurs en **distribution manuscrite** pour chaque réglage de la combinaison sans et avec prise en compte des scores selon le mécanisme (5.8), sur le **lot de courrier d’apprentissage**.

En conclusion, ce travail sur les scores, limité au courrier manuscrit, est très prometteur. Transposé au courrier dactylographié, il devrait permettre de franchir un nouveau seuil dans les performances.

5.3 Influence de la dépendance des LAP

Nous avons représenté, sur les figures 5.13 à 5.16, les nombres de bonnes reconnaissances et d'erreurs obtenus par les LAP Solystic et extérieur sur le lot de courrier d'apprentissage, pour les niveaux acheminement et distribution, et pour chaque type d'écriture.

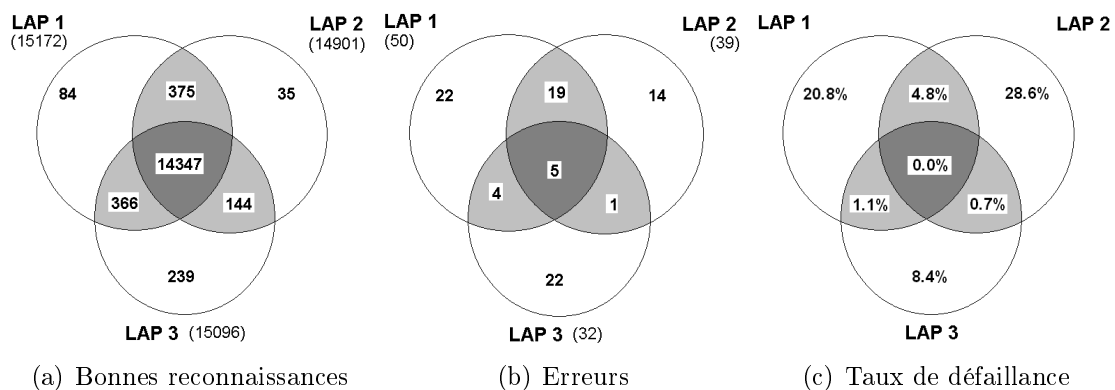


Figure 5.13 – Répartitions en **acheminement dactylographié**.

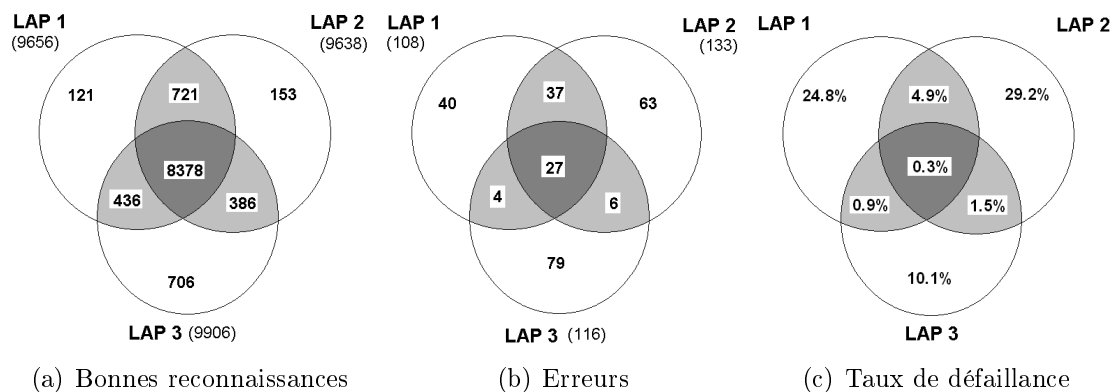


Figure 5.14 – Répartitions en **acheminement manuscrit**.

Par exemple, sur la figure 5.13(a) relative aux envois dactylographiés, nous pouvons observer que :

- 14347 décisions fournies conjointement par tous les LAP sur ces envois, sont correctes au niveau acheminement ;
- 375 décisions fournies conjointement par les LAP 1 et 2, le LAP 3 ayant fourni une autre décision, sont correctes au niveau acheminement ;

- de même, 144 décisions fournies conjointement par les LAP 2 et 3, alors que le LAP 1 fournissait une autre décision, sont correctes au niveau acheminement.
- En observant la figure 5.13(b) relative aux erreurs commises sur les envois dactylographiés, on remarque que :
- 5 décisions fournies conjointement par tous les LAP sur ces envois, sont erronées au niveau acheminement ;
 - 19 décisions communes aux LAP 1 et 2, le LAP 3 fournissant une autre décision, sont erronées au niveau acheminement ;
 - 1 décision fournie conjointement par les LAP 2 et 3, alors que le LAP 1 fournissait une autre décision, est erronée au niveau acheminement.

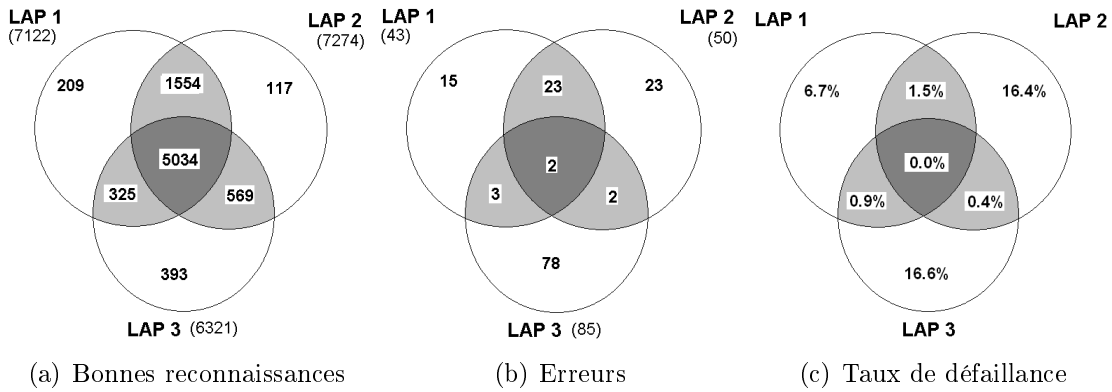


Figure 5.15 – Répartitions en **distribution dactylographiée**.

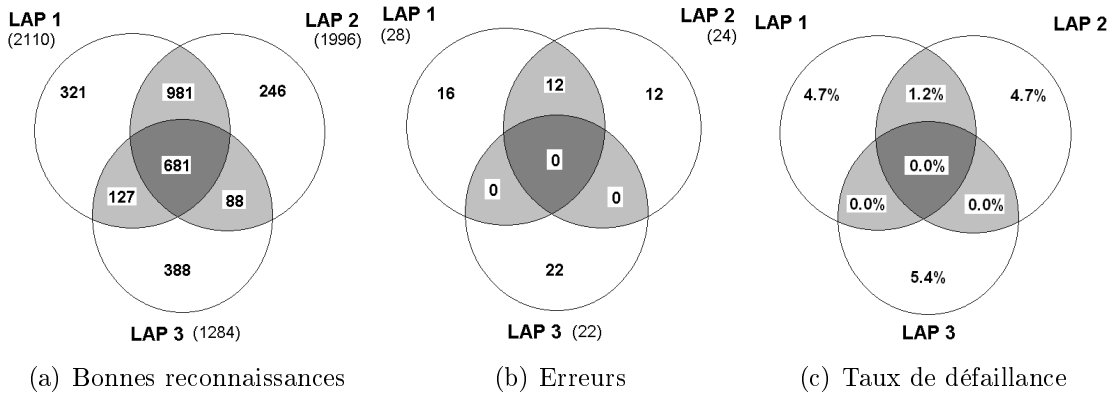


Figure 5.16 – Répartitions en **distribution manuscrite**.

Pour chacun des cas considérés sur les figures 5.13 à 5.16, la dernière figure représente le rapport entre le nombre de décisions erronées et la somme des décisions correctes et erronées, en fonction des LAP ayant fournies les décisions. Ce rapport est appelé *taux de défaillance*. Il correspond au complémentaire à un du taux de fiabilité défini au paragraphe 2.3.4. Ce taux représente la proportion d'erreur commise parmi les décisions exprimées à ce niveau. Ainsi, sur la figure 5.13(c), il est possible d'observer que :

- la proportion d’erreur commise, en acheminement sur les envois dactylographiés, pour les décisions fournies conjointement par tous les LAP est très faible : le taux de défaillance est de l’ordre de 0% ;
- en ce qui concerne les décisions fournies conjointement et uniquement par les LAP 1 et 2, le taux de défaillance est de l’ordre de 4.8% ;
- pour les décisions fournies conjointement et uniquement par les LAP 1 et 3 (resp. LAP 2 et 3), le taux de défaillance est égal à 1.1% (resp. 0.7%).

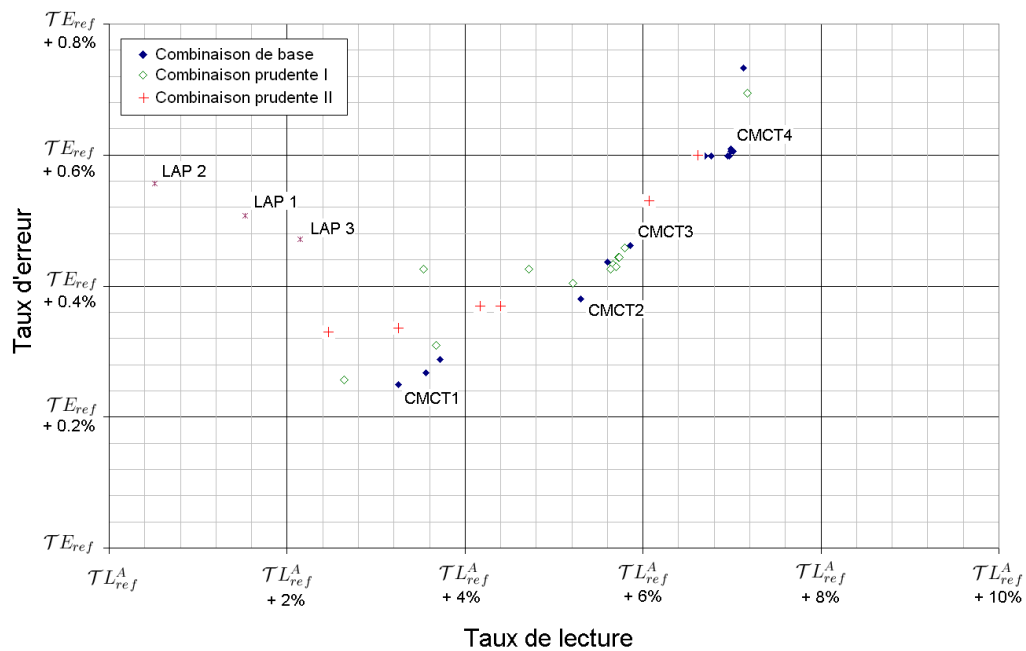
Les taux de défaillance des décisions fournies conjointement par le LAP extérieur et au moins un LAP Solystic sont donc toujours plus faibles que le taux de défaillance des décisions fournies conjointement et uniquement par les deux LAP Solystic. L’observation des taux de défaillance au niveau distribution et sur les envois manuscrits confirme cette remarque (cf figures 5.14(c), 5.15(c), et 5.16(c)). Ainsi, sur le lot de courrier d’apprentissage, une décision commune aux LAP 1 et 2 d’origine Solystic a plus de chance d’être une erreur qu’une décision commune au LAP 3 et à l’un des LAP Solystic. Ces observations s’expliquent par le fait que les deux LAP Solystic possèdent des algorithmes en commun, tandis que le LAP 3 d’origine extérieure possède des systèmes de reconnaissance et des méthodes de prise de décision différents de ceux mis en œuvre dans les LAP Solystic.

Afin de prendre en compte cette dépendance entre les LAP Solystic, deux combinaisons plus prudentes que la combinaison conjonctive employée dans le modèle base, ont été testées. Dans ces deux combinaisons, les informations fournies par les LAP Solystic sont combinées par la moyenne. La première combinaison prudente consiste à combiner conjonctivement cette moyenne avec la fonction de masse issue du LAP 3. Elle est donc définie par :

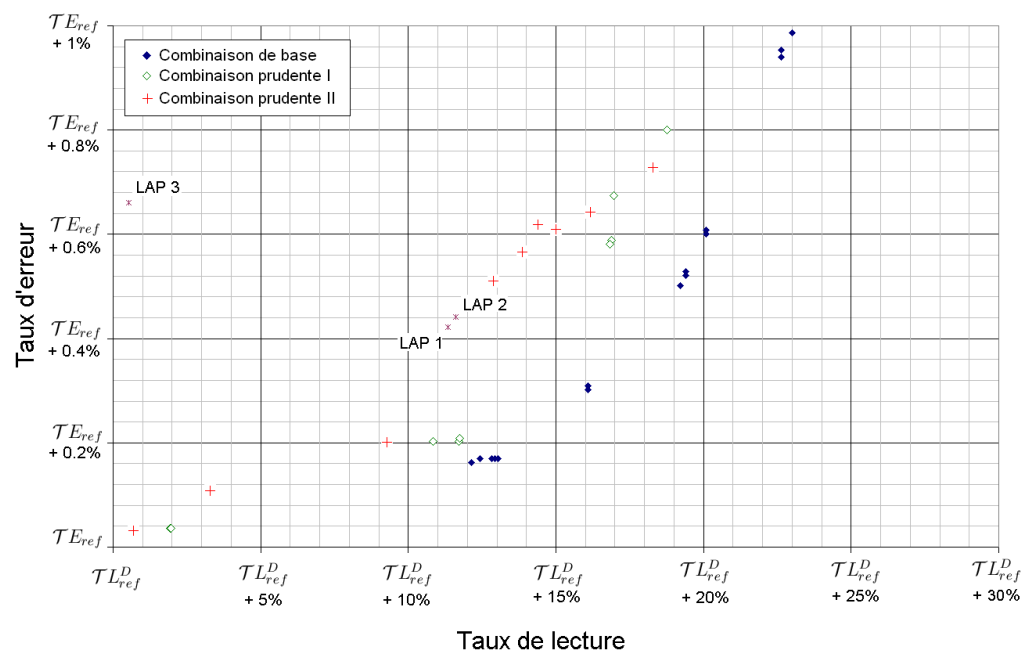
$$m = \frac{m_1 + m_2}{2} \odot m_3 . \quad (5.9)$$

La seconde combinaison prudente consiste à combiner cette moyenne avec la fonction de masse issue du LAP 3 par la combinaison de Dubois et Prade [37]. Au lieu de toujours transférer le produit des masses sur l’intersection des éléments focaux comme le fait la combinaison conjonctive (équation 2.21), cette combinaison consiste à transférer ce produit sur l’intersection si celle-ci est non vide, sinon sur l’union. Cette combinaison n’est pas associative, ainsi l’ordre dans lequel les sources sont combinées à une importance. Néanmoins, seulement deux informations sont considérées dans notre cas : une information Solystic et une information extérieure. Ainsi, cette question ne se pose pas.

Les performances de ces deux combinaisons sont illustrées sur la figure 5.17. Malgré quelques points intéressants en acheminement entre les points $CMCT_1$ et $CMCT_3$, les performances de ces combinaisons au niveau acheminement, et surtout au niveau distribution ne sont pas satisfaisantes. De nouvelles expérimentations demeurent à entreprendre. Par exemple, la combinaison prudente introduite récemment par Denœux [28] pourrait être employée. Ce travail est laissé en perspective.



(a) Performances au **niveau acheminement**.



(b) Performances au **niveau distribution**.

Figure 5.17 – Performances des LAP, de la combinaison de base et des deux combinaisons prudentes obtenues sur le **lot de courrier d'apprentissage**.

5.4 Prise en compte de décisions filtrées

Dans ce paragraphe, une perspective d'amélioration de cette fusion postale est exposée. Le modèle n'a pas encore été implémenté au jour de la rédaction de ce chapitre, néanmoins il est permis d'entrevoir le potentiel de la combinaison à court terme.

Lorsque le LAP effectue un rejet partiel ou complet, parmi toutes les décisions éliminées car jugées peu fiables, une première décision candidate pouvait se démarquer. Nous dirons que celle-ci a été *filtrée*. Lors de premiers tests, il a été effectivement vérifié qu'une grande partie de ces solutions étaient des erreurs. Cependant, ces premières solutions candidates pour les LAP Solystic pourraient s'avérer utiles en comptant sur la redondance des informations et sur l'utilisation conjointe des matrices de confusion et des scores. Par exemple, la sortie de ces décisions filtrées va permettre d'individualiser le traitement du saut persistant mentionné dans la remarque 5.1, comme l'illustre l'exemple suivant.

Exemple 5.1. *Supposons dans un premier cas les sorties suivantes des LAP :*

- le LAP 1 décide : « $A_i Rej$ », et a filtré « $A_i R_k$ »,
- le LAP 2 décide : « $A_i Rej$ », et a filtré « $A_i R_k$ »,
- le LAP 3 décide : « $A_i R_k$ »,

et, dans un deuxième cas les sorties suivantes :

- le LAP 1 décide : « $A_i Rej$ », et a filtré « $A_i R_{k'}$ »,
- le LAP 2 décide : « $A_i Rej$ », et n'a aucune solution distribution se démarquant,
- le LAP 3 décide : « $A_i R_k$ ».

Le premier cas est plus favorable que le second pour décider « $A_i R_k$ ».

À partir d'une matrice de confusion, une décision filtrée peut être convertie en une fonction de masse par la même affectation postale définie au paragraphe 3.2.3. En effet, en plus d'être caractérisée par sa catégorie et le type d'écriture du courrier traité, il suffit de remarquer qu'une décision peut aussi se distinguer en tant que décision filtrée ou non. Notons aussi qu'une décision filtrée peut être accompagnée d'un score.

Remarque 5.2 (Émergence). *Dans les modèles exposés avant ce paragraphe, il ne peut être fourni qu'une décision de niveau identique ou moins fin que le plus fin des niveaux des décisions proposées par les LAP. Avec ces décisions filtrées, une décision de niveau strictement plus fin peut être envisagée. Ainsi, la combinaison pourrait décider des adresses postales complètes non décidées par les LAP à fusionner. La combinaison prendrait alors tout son sens : la véritable adresse d'un envoi postal ne pourrait être obtenue que par une action collective des LAP. Individuellement, aucun LAP ne pourrait trouver la solution.*

5.5 Conclusion

Dans ce chapitre, les derniers résultats obtenus lors de notre travail de thèse ont été exposés. L'apport des scores sur les envois manuscrits a permis d'améliorer encore les performances obtenues par le modèle de base. Ces performances très

encourageantes dépassent nettement les performances individuelles des LAP et des combinaisons à base de votes. En incluant des scores associés au courrier dactylographié et en considérant les décisions filtrées, les performances de ce modèle de combinaison devraient à nouveau franchir un palier.

Conclusion générale et perspectives

Dans ce mémoire, un problème de fusion de décisions postales a été présenté. Une des caractéristiques de ce problème réside dans l'organisation hiérarchique des décisions, chaque LAP pouvant fournir une décision d'un niveau quelconque dans la hiérarchie. Plusieurs modèles de fusion de décisions postales ont été construits dans le cadre de la théorie des fonctions de croyance.

Un premier modèle s'appuie sur une affectation des fonctions de masses construites à partir d'une matrice de confusion et de la hiérarchie des décisions postales. La méthode de prise de décision adoptée permet un ajustement de la combinaison aux contraintes de taux de lecture et taux d'erreur attendues par l'utilisateur. Cette combinaison peut s'appliquer à tout type de LAP d'origine Solystic ou extérieure, et fournit d'ores et déjà des performances intéressantes.

Sur le plan théorique, de nouveaux outils de manipulation de fonction de croyance ont été développés afin d'exploiter plus finement des informations sur la fiabilité des sources à combiner. Par l'affaiblissement contextuel, la dépendance au contexte de la fiabilité peut être prise en compte. Par le mécanisme général de correction proposé, différents états de fiabilité peuvent être modélisés, ce qui permet de développer des stratégies de correction autres qu'un affaiblissement.

Ces deux mécanismes apportent leur contribution à l'amélioration du premier modèle. Nous avons montré comment l'information fournie par un LAP ne travaillant pas sur l'espace complet des décisions, pouvait être prise en compte par un affaiblissement contextuel. L'utilisation d'un mécanisme général de correction permettant de renforcer ou d'affaiblir l'information fournie par un LAP en fonction du score accompagnant sa décision, a été particulièrement bénéfique.

En perspective à court terme, l'ajout de scores au regard du courrier dactylographié et la sortie de décisions filtrées mentionnée dans le dernier chapitre, devraient confirmer les bonnes performances de ce modèle de combinaison.

À plus long terme, les perspectives demeurent nombreuses. Parmi les nouvelles informations pouvant être disponibles, citons les décisions parentes et filtrées, associées à un score. De même, la sortie de décisions complémentaires arrivant en deuxième ou troisième position serait possible. Plusieurs décisions différentes seraient alors issues d'un LAP. Nous pensons que ces informations peuvent être représentées, comme n'importe quelle information, par des fonctions de croyance. Plusieurs modèles peuvent être construits en fonction de l'interprétation donnée à ces informations. Un premier modèle consisterait à étendre l'ancien mécanisme de correction de façon à considérer plusieurs scores. Une seconde approche consisterait à convertir directement les scores en une nouvelle fonction de masse, qui viendrait en complément de l'affectation postale basée sur la matrice de confusion du LAP.

Une méthode simple consisterait à associer à chaque décision une masse égale à son score, puis à normaliser.

D'autre part, une autre stratégie pourrait être envisagée pour initialiser les masses d'une fonction de croyance associée à la décision d'un LAP. Outre l'utilisation des matrices de confusion et des scores, un ensemble de courriers synthétiques pourrait être créé. La soumission de ce lot de courrier permettrait, par exemple, de mesurer la capacité d'un LAP à détecter différentes polices de caractères, ainsi que la bonne adresse en présence d'une adresse expéditeur ou de ratures, et à tester le comportement du LAP sur des grandes catégories de courrier (enveloppe à fenêtre, carte postale, courrier de grandes administrations) ou en présence de certaines incohérences sur un envoi.

Plus généralement, la question de savoir quels autres éléments sont nécessaires à la combinaison, demeure à explorer. D'un côté, nous serions tenté de toujours demander plus d'informations aux LAP pour affiner de plus en plus les résultats de la combinaison. Mais, d'un autre côté, nous souhaiterions que la combinaison demeure un module indépendant de l'ensemble des LAP disponibles. Dans cette application postale, l'adresse à identifier est issue d'écritures sur une enveloppe. Les mots sur cette enveloppe jouent un rôle important. Or, si les intitulés de mot ou de phrase reconnus isolément et partiellement deviennent accessibles à la combinaison, leurs interprétations risquent de transformer la combinaison en un nouveau LAP Solystic, ce qui est exclu. Ce dilemme pourrait être résolu en sélectionnant uniquement des éléments de bas niveau jugés pertinents, leur identification demeurant à explorer.

Enfin, au regard de la manière simple avec laquelle chaque information a pu être représentée par une fonction de croyance, des performances obtenues dans ce cadre, qui dépassent ce qui était précédemment réalisée, et des modèles non encore implémentés au jour de la rédaction de ce mémoire, ce travail de fusion postale crédale est véritablement très prometteur.

Bibliographie

- [1] A. Appriou. Perspectives liées à la fusion de données. *Science et Défense* 90, 1990.
- [2] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, 11 :27–40, 1991.
- [3] A. Appriou. Multisensor signal processing in the framework of the theory of evidence. *NATO/RTA, SCI Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, RTO EN-7 :(5–1)–(5–31), 1999.
- [4] N.O. Attoh-Okine. Aggregating evidence in pavement management decision-making using belief functions and qualitative markov tree. *IEEE Transactions on Systems, Man and Cybernetics C*, 32(3) :243–251, 2002.
- [5] A. Ayoun et Ph. Smets. Data association in multi-target detection using the transferable belief model. *International Journal of Intelligent Systems*, 16(10) :1167–1182, 2001.
- [6] J.A. Barnett. Computational methods for a mathematical theory of evidence. Dans *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81)*, pages 868–875, 1981.
- [7] A. Bastiere. Methods for multisensor classification of airborne targets integrating evidence theory. *Aerospace Science and Technology*, 2(6) :401–411, 1998.
- [8] A. Ben Yaghlane, T. Dencœux, et K. Mellouli. Elicitation of expert opinions for constructing belief functions. Dans *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2006*, pages 403–411. Paris, France, July 2-7 2006.
- [9] B. Benyoub. Une application industrielle de reconnaissance d'adresses. Dans *Colloque National sur l'Écrit et le Document*, pages 93–100. Nantes, France, Juillet 1996.

- [10] H. Bersini, T. Dencœux, D. Dubois, et H. Prade. In memoriam : Philippe smets (1938-2005). *Information Fusion*, 7(2) :240–244, June 2006.
- [11] I. Bloch. Incertitude, imprécision et additivité en fusion de données : point de vue historique. *Traitement du signal*, 13(4) :267–288, 1996.
- [12] I. Bloch. Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, 17(8) :905–919, 1996.
- [13] I. Bloch. *Fusion d'informations en traitement du signal et des images*. Hermès, Paris, France, 2003.
- [14] B. Bouchon-Meunier. *La logique floue et ses applications*. Addison-Wesley France, Paris, France, 1995.
- [15] B.R. Cobb et P.P. Shenoy. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3) :314–330, April 2006.
- [16] M. Daniel. Probabilistic transformations of belief functions. Dans Lluís Godo (Ed), *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 539–551. Springer Verlag, Barcelona, Spain, July 6-8 2005.
- [17] B. de Finetti. *Teoria della probabilità*. 2 volumi, G. Einaudi Editore, Torino, Réimprimé en anglais en 1974 par John Wiley and Sons, New York, 1970.
- [18] G. De Leo, M. Vicenzi, et C. Franzone. ELSAG SPA. European patent : Mail recognition method. Number : EP 1 594 077 A2, November 2005.
- [19] M.H. DeGroot. *Optimal Statistical decisions*. McGraw-Hill, New York, 1970.
- [20] A.P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Statistics*, 38 :325–339, 1967.
- [21] A.P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society*, 30 :205–247, 1968.
- [22] A.P. Dempster. A class of random convex polytopes. *Annals of Mathematical Statistics*, 43 :250–272, 1972.
- [23] T. Dencœux. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5) :804–813, 1995.
- [24] T. Dencœux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7) :1095–1107, 1997.

- [25] T. Denœux. Application du modèle des croyances transférables en reconnaissance de formes. *Traitement du Signal*, 14(5) :443–451, 1998.
- [26] T. Denœux. A neural network classifier based on dempster-shafer theory. *IEEE transactions on Systems, Man and Cybernetics A*, 30(2) :131–150, 2000.
- [27] T. Denœux. Théorie des fonctions de croyance et classification. Séminaire thème ASTRID, Heudiasyc, Compiègne, Juin 2004.
- [28] T. Denœux. The cautious rule of combination for belief functions and some extensions. Dans *Proceedings of the 9th International Conference on Information Fusion, FUSION'2006*. Florence, Italy, July 10-14 2006.
- [29] T. Denœux et Ph. Smets. Classification using belief functions : the relationship between the case-based and model-based approaches. *IEEE transactions on Systems, Man and Cybernetics B*, 36(6) :1395–1406, 2006.
- [30] S. Démotier, W. Schön, et T. Denœux. Risk assessment based on weak information using belief functions : A case study in water treatment. *IEEE Transactions on Systems, Man and Cybernetics C*, 36(3) :382–396, 2006.
- [31] D. Dubois et H. Prade. On several representations of an uncertain body of evidence. Dans M.M. Gupta et E. Sanchez (Eds), *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, 1982.
- [32] D. Dubois et H. Prade. A note on measures of specificity for fuzzy sets. *International Journal of General Systems*, 10(4) :279–283, 1985.
- [33] D. Dubois et H. Prade. *Théorie des possibilités. Application à la représentation des connaissances en informatique*. Masson, Paris, France, 1985.
- [34] D. Dubois et H. Prade. On the unicity of Dempster rule of combination. *International Journal of Intelligent Systems*, 1 :133–142, 1986.
- [35] D. Dubois et H. Prade. A set-theoretic view of belief functions : logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12 :193–226, 1986.
- [36] D. Dubois et H. Prade. The principle of minimum specificity as a basis for evidential reasoning. Dans B. Bouchon et R.R. Yager (Eds), *Uncertainty in Knowledge-Based Systems*, pages 75–84. Springer Verlag, 1987.
- [37] D. Dubois et H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4 :244–264, 1988.
- [38] D. Dubois et H. Prade. Focusing vs. belief revision : A fundamental distinction when dealing with generic knowledge. Dans A. Nonnengart

- D. Gabbay, R. Kruse et H.J. Ohlbach (Eds), *Proceedings of the 1st International Joint Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU/FAPR'97*, pages 96–107. Springer Verlag, Berlin, 1997.
- [39] D. Dubois, H. Prade, et Ph. Smets. Representing partial ignorance. *IEEE Transactions on Systems, Man and Cybernetics*, 26(3) :361–377, 1996.
- [40] Z. Elouedi, K. Mellouli, et Ph. Smets. Assessing sensor reliability for multisensor data fusion with the transferable belief model. *IEEE Transactions on Systems, Man and Cybernetics B*, 34 :782–787, 2004.
- [41] S. Fabre, A. Appriou, et X. Briottet. Presentation and description of two classification methods using data fusion based on sensor management. *Information Fusion*, 2 :49–71, 2001.
- [42] M. Fisher. Siemens AG. International patent : System and method for smart polling. Number : WO 2005/050545 A1, June 2005.
- [43] G. Giacinto, R. Paolucci, et F. Roli. Application of neural networks and statistical pattern recognition algorithms to earthquake risk evaluation. *Pattern Recognition Letters*, 18(11-13) :1353–1362, 1997.
- [44] J. Gordon et E.H. Shortliffe. The Dempster-Shafer theory of evidence. Dans B.G. Buchanan et E.H. Shortliffe (Eds), *Rule-based expert systems : The MYCIN experiments of the Stanford Heuristic Programming Project*, chapitre 13, pages 272–292. Addison-Wesley, Reading, MA, 1984.
- [45] D. Gruyer. *Étude du traitement de données imparfaites pour le suivi multi-objets : Application. aux situations routières*. Thèse de doctorat, Université de Technologie de Compiègne, 1999.
- [46] R. Haenni. Ignoring ignorance is ignorant. Working paper of the philosophy, probability, and modeling research group, University of Konstanz, 2003.
- [47] R. Haenni, J. Kohlas, et N. Lehmann. Probabilistic argumentation systems. Dans J. Kohlas et S. Moral (Eds), *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 5, pages 221–288. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [48] P. Hájek. Deriving Dempster's rule. Dans *Proceedings of the 4th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'92*, pages 73–75, 1992.
- [49] Z. Hammal, A. Caplier, et M. Rombaut. A fusion process based on belief theory for classification of facial basic emotions. Dans *Proceedings of the 8th International Conference on Information Fusion, FUSION'2005*. Philadelphie, E-U, July 25-29, paper C1-3, 2005.

- [50] Y.-T. Hsia. Characterizing belief with minimum commitment. Dans *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI'91)*, pages 1184–1189, 1991.
- [51] H. Katsuno et A. Mendelzon. On the difference between updating a knowledge base and revising it. Dans J.F. Allen, R. Fikes, et E. Sandewall (Eds), *Principles of Knowledge Representation and Reasoning (KR'91)*, pages 387–394. Morgan Kaufmann, San Mateo, CA, 1991.
- [52] J. Kittler, M. Hatef, R.P.W. Duin, et J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3) :226–239, March 1998.
- [53] F. Klawonn et E. Schwecke. On the axiomatic justification of Dempster's rule of combination. *International Journal of Intelligent Systems*, 7 :469–478, 1992.
- [54] F. Klawonn et Ph. Smets. The dynamic of belief in the transferable belief model and specialization-generalization matrices. Dans B. D'Ambrosio D. Dubois, M. P. Wellman et Ph Smets (Eds), *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, pages 130–137. San Mateo, CA, 1992.
- [55] G.J. Klir et M.J. Wiermann. *Uncertainty-Based Information. Elements of Generalized Information Theory*. Physica-Verlag, Heidelberg, 1998.
- [56] J. Kohlas et P.-A. Monney. *A Mathematical Theory of Hints. An Approach to the Dempster-Shafer Theory of Evidence*, volume 425. Springer-Verlag, Berlin, 1995.
- [57] R. Kruse et E. Schwecke. Specialization : a new concept for uncertainty handling with belief functions. *International Journal of General Systems*, 18 :49–60, 1990.
- [58] L. Lam et C.Y. Suen. Application of majority voting to pattern recognition : An analysis of its behaviour and performance. *IEEE Transactions on systems, Man and Cybernetics*, 27 :553–568, 1997.
- [59] S. Le Hégarat-Masclé, I. Bloch, et D. Vidal-Madjar. Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover. *Pattern recognition*, 31(11) :1811–1823, November 1998.
- [60] E. Lefevre. *Fusion adaptée d'informations conflictuelles dans le cadre de la théorie de l'évidence, application au diagnostic médicale*. Thèse de doctorat, Institut National des Sciences Appliquées de Rouen, 2001.
- [61] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, MA, 1980.

- [62] E. Mandler et J. Schürmann. Combining the classification results of independent classifiers based on the dempster-shafer theory of evidence. Dans E. S. Gelsema et L. N. Kanal (Eds), *Pattern Recognition and Artificial Intelligence*, pages 381–393. North-Holland, Amsterdam, 1988.
- [63] D. Mercier, G. Cron, et B. Benyoub. Méthode pour fusionner des OCR postaux exploitant des fonctions de croyance. Dépôt INPI n° 06 509 10, 17 Mars 2006.
- [64] D. Mercier, G. Cron, T. Dencœux, et M. Masson. Fusion de décisions postales dans le cadre du modèle des croyances transférables. *Traitement du signal, numéro spécial : « Théorie des fonctions de croyance : de nouveaux horizons pour l'aide à la décision »*, à paraître.
- [65] D. Mercier, G. Cron, T. Dencœux, et M. Masson. Fusion of multi-level decision systems using the transferable belief model. Dans *Proceedings of the 8th International Conference on Information Fusion, FUSION'2005*. Philadelphie, E-U, paper C8-2, July 25-29, 2005.
- [66] D. Mercier, G. Cron, T. Dencœux, et M. Masson. Vers un modèle de fusion de décisions de lecteurs d'adresses postales basé sur la théorie des fonctions de croyance. Dans L. Likforman-Sulem (Ed), *Actes du 9e Colloque International Francophone sur l'Écrit et le Document, CIFED'2006*, pages 79–84. Fribourg, Suisse, 18-21 Septembre 2006.
- [67] D. Mercier, G. Cron, T. Dencœux, et M. Masson. Une approche globale de fusion d'adresses postales basée sur la théorie des fonctions de croyance. Dans *Actes des 15e rencontres francophones sur la Logique Floue et ses Applications, LFA'2006*. Toulouse, 19-20 Octobre 2006, à paraître.
- [68] D. Mercier, T. Dencœux, et M. Masson. General correction mechanisms for weakening or reinforcing belief functions. Dans *Proceedings of the 9th International Conference on Information Fusion, FUSION'2006*. Florence, Italy, paper 146, July 10-14 2006.
- [69] D. Mercier, T. Dencœux, et M. Masson. Refined sensor tuning in the belief function framework using contextual discounting. Dans *Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2006*, pages 1443–1450. Paris, France, July 2-7 2006.
- [70] D. Mercier, B. Quost, et T. Dencœux. Contextual discounting of belief functions. Dans Lluís Godo (Ed), *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 552–562. Springer Verlag, Barcelona, Spain, July 6-8 2005.

- [71] D. Mercier, B. Quost, et T. Dencœux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, To appear in 2007.
- [72] N. Milisavljević et I. Bloch. Sensor fusion in anti-personnel mine detection using a two-level belief function model. *IEEE Transaction on Systems, Man and Cybernetics C*, 33(2) :269–283, 2003.
- [73] R.R. Murphy. Dempster-Shafer theory for sensor fusion in autonomous mobile robots. *IEEE Transactions on Robotics and Automation*, 14(2) :197–206, 1998.
- [74] F.P. Ramsey. Truth and probability. *The Foundations of Mathematics : Collected Papers of Frank P. Ramsey*, pages 156–198, 1931.
- [75] B. Ristic et Ph Smets. Recursive classification of multiple objects using discordant and non-specific data. *Submitted*, 2004.
- [76] B. Ristic et Ph Smets. Target identification using belief functions and implication rules. *IEEE Transactions on Aerospace and Electronic Systems*, 41(3) :1097–1103, 2005.
- [77] G. Rogova. Combining the results of several neural network classifiers. *Neural Networks*, 7(5) :777–781, 1994.
- [78] M. Rombaut. Decision in multi-obstacle matching process using theory of belief. Dans *Proceedings of the International Conference on Advances in Vehicle Control and Safety, AVCS'98*, pages 63–68. Amiens, July 1998.
- [79] M. Rombaut. Fusion : état de l'art et perspectives. Rapport DGA DSP 99.60.078, LIS UMR 5083 - INPG, 47 pages, 22 octobre 2001.
- [80] F. Rottensteiner, J. Trinder, S. Clode, et K. Kubik. Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*, 6(4) :283–300, 2005.
- [81] C. Royère. *Théorie de l'évidence en détection d'obstacles*. Thèse de doctorat, Université de Technologie de Compiègne, 2002.
- [82] C. Royère, D. Gruyer, et V. Cherfaoui. Data association with believe theory. Dans *Proceedings of the 3rd International Conference on Information Fusion, FUSION'2000*, pages TuD2–9. Paris, France, July 10-13 2000.
- [83] L.J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.
- [84] J. Schubert. On nonspecific evidence. *International Journal of Intelligent Systems*, 8(6) :711–725, 1993.

- [85] J. Schubert. *Cluster-based Specification Techniques in Dempster-Shafer Theory for an Evidential Intelligence Analysis of Multiple Target Tracks*. Thèse de doctorat, Royal Institute of Technology, Stockholm, 1994.
- [86] J. Schubert. Managing inconsistent intelligence. Dans *Proceedings of the 3rd International Conference on Information Fusion, FUSION'2000*, pages TuB4–10. Paris, France, July 10-13 2000.
- [87] K. Sentz et S. Ferson. Combination of evidence in Dempster-Shafer theory. Rapport SAND 2002-0835, Sandia National Laboratories, 96 pages, 2002.
- [88] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
- [89] G. Shafer. Non-additive probabilities in the works of Bernoulli and Lambert. *Archives for the History of Exact Sciences*, 19 :309–370, 1978.
- [90] G. Shafer. The combinaison of evidence. *International Journal of Intelligent Systems*, 1 :155–179, 1986.
- [91] G. Shafer. The early development of mathematical probability. Dans I. Grattan-Guinness (Ed), *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, pages 1293–1302. Routledge, London, 1993.
- [92] G. Shafer. The significance of Jacob Bernoulli's ars conjectandi for the philosophy of probability today. *Journal of Econometrics*, 75 :15–32, 1996.
- [93] P.P. Shenoy. A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3(2) :383–411, 1989.
- [94] Ph. Smets. *Un modèle mathématico-statistique simulant le processus du diagnostic médical*. Thèse de doctorat, Université Libre de Bruxelles, 1978.
- [95] Ph. Smets. The combination of evidence in the transferable belief model. *IEEE Pattern analysis and Machine Intelligence*, 12 :447–458, 1990.
- [96] Ph. Smets. The concept of distinct evidence. Dans *Proceedings of the 4th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'92*, pages 789–794. Palma de Mallorca, 1992.
- [97] Ph. Smets. The nature of the unnormalized beliefs encountered in the transferable belief model. Dans B. D'Ambrosio D. Dubois, M. P. Wellman et Ph Smets (Eds), *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, pages 292–297. San Mateo, CA, 1992.
- [98] Ph. Smets. Belief functions : the disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9 :1–35, 1993.

- [99] Ph. Smets. What is Dempster-Shafer's model ? Dans R. R. Yager, J. Kacprzyk, et M. Fedrizzi (Eds), *Advances in the Dempster-Shafer theory of evidence*, pages 5–34. Wiley, New-York, 1994.
- [100] Ph. Smets. The axiomatic justification of the transferable belief model. Technical Report TR/IRIDIA/95-8, IRIDIA, Bruxelles, 1995.
- [101] Ph. Smets. The canonical decomposition of a weighted belief. Dans *Proceedings of the 14th International Joint Conference on Artificial Intelligence, (IJCAI'95)*, pages 1896–1901. Morgan Kaufmann Publishers, San Mateo, CA, 1995.
- [102] Ph. Smets. The α -junctions : the commutative combination operators applicable to belief functions. Dans A. Nonnengart D. Gabbay, R. Kruse et H.J. Ohlbach (Eds), *Proceedings of the 1st International Joint Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU/FAPR'97*, pages 131–153. Springer Verlag, Berlin, 1997.
- [103] Ph. Smets. The Transferable Belief Model for quantified belief representation. Dans D. M. Gabbay et Ph. Smets (Eds), *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 267–301. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
- [104] Ph. Smets. Practical uses of belief functions. Dans K. B. Laskey et H. Prade (Eds), *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 612–621. Stockholm, Sweden, 1999.
- [105] Ph. Smets. Beliefs functions and the transferable belief model. *Contribution à la documentation en ligne du site internet de la société SIPTA (Society for Imprecise Probability Theory and Applications)*, 2000.
- [106] Ph. Smets. Data fusion in the transferable belief model. Dans *Proceedings of the 3rd International Conference on Information Fusion, FUSION'2000*, pages PS21–33. Paris, France, July 10-13 2000.
- [107] Ph. Smets. The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31(1–2) :1–30, 2002.
- [108] Ph. Smets. Belief functions on real numbers. *International Journal of Approximate Reasoning*, 40(3) :181–223, 2005.
- [109] Ph. Smets. Decision making in the TBM : the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2) :133–147, 2005.
- [110] Ph. Smets. Analyzing the combination of conflicting belief functions. *Disponible en ligne http://iridia.ulb.ac.be/psmets/Combi_Confl.pdf*. À paraître dans *Information Fusion*, 2006.

- [111] Ph. Smets et R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66 :191–243, 1994.
- [112] C.A.B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society*, B23 :1–37, 1961.
- [113] L. Sombe. A glance at revision an updating in knowledge bases. *International Journal of Intelligent Systems*, 9 :1–28, 1994.
- [114] Traitement du signal (Eds.). Numéro special : Fusion de données. *Traitement du signal*, 14(5), 1997.
- [115] F. Tupin. *Reconnaissance des formes et analyse de scènes en imagerie radar à ouverture synthétique*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, 1997.
- [116] F. Tupin, I. Bloch, et H. Maître. A First Step Towards Automatic Interpretation of SAR Images using Evidential Fusion of Several Structure Detectors. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3) :1327–1343, 1999.
- [117] P. Vannoorenberghe. Un état de l’art sur les fonctions de croyance appliquées au traitement de l’information. *Revue I3*, 3(2) :9–45, 2003.
- [118] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [119] P. Walley et S. Moral. Upper probabilities based only on the likelihood function. *Journal of the Royal Statistical Society, Series B*, 61 :831–847, 1999.
- [120] L. Xu, A. Krzyzak, et C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3) :418–435, 1992.
- [121] R.R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9(4) :249–260, 1983.
- [122] R.R. Yager. The entailment principle for Dempster-Shafer granules. *International Journal of Intelligent Systems*, 1 :247–262, 1986.
- [123] R.R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41 :93–137, 1987.
- [124] L.A. Zadeh. Fuzzy Sets. *Information and Control*, 3(8) :338–353, 1965.
- [125] L.A. Zadeh. Book review : A mathematical theory of evidence. *Artificial Intelligence Magazine*, 5 :81–83, 1984.

- [126] M. Zaffalon et G. de Cooman. Editorial : Imprecise probability perspectives on artificial intelligence. *Annals of Mathematics and Artificial Intelligence*, 45(1-2) :1–4, 2005.
- [127] H. Zhu et O. Basir. Extended discounting scheme for evidential reasoning as applied to MS lesion detection. Dans Per Svensson et Johan Schubert (Eds), *Proceedings of the 7th International Conference on Information Fusion, FUSION'2004*, volume I, pages 280–287. International Society of Information Fusion, Mountain View, CA, June 2004.
- [128] L. M. Zouhal et T. Dencœux. An evidence-theoretic k-nn rule with parameter optimization. *IEEE Transactions on Systems, Man and Cybernetics C*, 28(2) :263–271, 1998.

Preuves des énoncés du chapitre 4

A.1 Preuves des propositions 4.1 et 4.2

Le déconditionnement de $m_{Ag}^{\mathcal{R}}[\omega_k]$ est donné par :

$$m_{Ag}^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(\{\omega_k\} \times \{R\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) = \beta_k, \quad (\text{A.1})$$

$$m_{Ag}^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(\Omega \times \mathcal{R}) = \alpha_k. \quad (\text{A.2})$$

Notons $m_r^{\Omega\times\mathcal{R}}$ la combinaison conjonctive des $m_{Ag}^{\mathcal{R}}[\omega_k]^{\uparrow\Omega\times\mathcal{R}}$, $k = 1, \dots, K$. En utilisant l'égalité suivante, pour tout $k \neq \ell$:

$$(\{\omega_k\} \times \{R\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) \cap (\{\omega_\ell\} \times \{R\} \cup \overline{\{\omega_\ell\}} \times \mathcal{R}) = \{\omega_k, \omega_\ell\} \times \{R\} \cup \overline{\{\omega_k, \omega_\ell\}} \times \mathcal{R},$$

l'expression de $m_r^{\Omega\times\mathcal{R}}$ peut être aisément obtenue :

$$m_r^{\Omega\times\mathcal{R}}(C \times \{R\} \cup \overline{C} \times \mathcal{R}) = \begin{cases} \prod_{\substack{\omega_k \in \overline{C} \\ k=1 \\ K}} \alpha_k \prod_{\omega_\ell \in C} \beta_\ell & \text{si } C \neq \emptyset \text{ et } C \neq \Omega, \\ \prod_{\substack{k=1 \\ K}} \alpha_k & \text{si } C = \emptyset, \\ \prod_{\ell=1} \beta_\ell & \text{si } C = \Omega. \end{cases} \quad (\text{A.3})$$

En échangeant les rôles de C et \overline{C} , nous obtenons :

$$m_r^{\Omega\times\mathcal{R}}(\overline{C} \times \{R\} \cup C \times \mathcal{R}) = \begin{cases} \prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell & \text{si } C \neq \emptyset \text{ et } C \neq \Omega, \\ \prod_{\substack{k=1 \\ K}} \alpha_k & \text{si } C = \Omega, \\ \prod_{\ell=1} \beta_\ell & \text{si } C = \emptyset, \end{cases} \quad (\text{A.4})$$

ce qui peut se noter plus simplement :

$$m_r^{\Omega\times\mathcal{R}}(\overline{C} \times \{R\} \cup C \times \mathcal{R}) = \prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell, \quad \forall C \subseteq \Omega. \quad (\text{A.5})$$

Cette fonction de masse doit être combinée avec $m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}}$:

$$\alpha m_{Ag}^\Omega = \left(m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}} \odot m_r^{\Omega \times \mathcal{R}} \right)^{\downarrow\Omega}. \quad (\text{A.6})$$

Les éléments focaux des fonctions de masse $m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}}$ et $m_r^{\Omega \times \mathcal{R}}$ sont, respectivement, de la forme $B \times \{R\} \cup \Omega \times \{NR\}$ et $\bar{C} \times \{R\} \cup C \times \mathcal{R}$, avec $B, C \subseteq \Omega$. L'intersection de ces éléments focaux est de la forme :

$$(\bar{C} \times \{R\} \cup C \times \mathcal{R}) \cap (B \times \{R\} \cup \Omega \times \{NR\}) = B \times \{R\} \cup C \times \{NR\},$$

et, cette intersection ne peut être obtenue que pour un choix particulier de B et C . Ainsi :

$$m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}} \odot m_r^{\Omega \times \mathcal{R}}(B \times \{R\} \cup C \times \{NR\}) = \left[\prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \bar{C}} \beta_\ell \right] m_S^\Omega(B). \quad (\text{A.7})$$

En projetant ce résultat sur Ω , nous obtenons :

$$\alpha m^\Omega(A) = \sum_{B \cup C = A} \left[\prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \bar{C}} \beta_\ell \right] m_S^\Omega(B), \quad \forall A \subseteq \Omega, \quad (\text{A.8})$$

ce qui peut aussi s'écrire :

$$\alpha m^\Omega(A) = \sum_{B \subseteq A} \alpha G(A, B) m_S^\Omega(B), \quad \forall A \subseteq \Omega, \quad (\text{A.9})$$

avec :

$$\alpha G(A, B) = \sum_{C: B \cup C = A} \prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \bar{C}} \beta_\ell, \quad \forall B \subseteq A \subseteq \Omega, \quad (\text{A.10})$$

et $\alpha G(A, B) = 0$ si $B \not\subseteq A$. Maintenant, nous avons

$$\begin{aligned} B \cup C = A &\Leftrightarrow \exists D \subseteq B, C = (A \setminus B) \cup D \\ &\Leftrightarrow \exists D \subseteq B, \bar{C} = \bar{A} \cup (B \setminus D), \end{aligned}$$

et, par conséquent, pour tout A et B tels que $B \subseteq A$:

$$\begin{aligned} \alpha G(A, B) &= \sum_{D \subseteq B} \prod_{\omega_k \in (A \setminus B) \cup D} \alpha_k \prod_{\omega_\ell \in \bar{A} \cup (B \setminus D)} \beta_\ell \\ &= \prod_{\omega_k \in A \setminus B} \alpha_k \prod_{\omega_\ell \in \bar{A}} \beta_\ell \sum_{D \subseteq B} \prod_{\omega_k \in B \setminus D} \beta_k \prod_{\omega_\ell \in D} \alpha_\ell \\ &= \prod_{\omega_k \in A \setminus B} \alpha_k \prod_{\omega_\ell \in \bar{A}} \beta_\ell \prod_{\omega_k \in B} (\alpha_k + \beta_k) \\ &= \prod_{\omega_k \in A \setminus B} \alpha_k \prod_{\omega_\ell \in \bar{A}} \beta_\ell, \end{aligned}$$

ce qui conclut la preuve de la proposition 4.1.

La proposition 4.2 est évidente à partir de l'équation A.8.

□

A.2 Modification des croyances sur la fiabilité

Ce paragraphe aborde les solutions obtenues en modifiant l'expression des croyances de l'agent sur la fiabilité de la source, équation (4.22).

Cas 1 : Supposons que l'équation 4.22 soit remplacée par :

$$\begin{cases} m_{Ag}^{\mathcal{R}}[\omega_k](\{R\}) &= \beta_k, \\ m_{Ag}^{\mathcal{R}}[\omega_k](\{NR\}) &= \alpha_k. \end{cases} \quad (\text{A.11})$$

Dans ce cas, le déconditionnement de $m_{Ag}^{\mathcal{R}}[\omega_k]$ est :

$$m_{Ag}^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(\{\omega_k\} \times \{R\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) = \beta_k, \quad (\text{A.12})$$

$$m_{Ag}^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(\{\omega_k\} \times \{NR\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) = \alpha_k. \quad (\text{A.13})$$

Soit $m_r^{\Omega\times\mathcal{R}}$ la combinaison conjonctive des $m_{Ag}^{\mathcal{R}}[\omega_k]^{\uparrow\Omega\times\mathcal{R}}$, $k = 1, \dots, K$. En utilisant les égalités suivantes, pour tout $k \neq \ell$:

$$\begin{aligned} (\{\omega_k\} \times \{R\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) \cap (\{\omega_\ell\} \times \{R\} \cup \overline{\{\omega_\ell\}} \times \mathcal{R}) = \\ \{\omega_k, \omega_\ell\} \times \{R\} \cup \overline{\{\omega_k, \omega_\ell\}} \times \mathcal{R}, \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} (\{\omega_k\} \times \{R\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) \cap (\{\omega_\ell\} \times \{NR\} \cup \overline{\{\omega_\ell\}} \times \mathcal{R}) = \\ \{\omega_k\} \times \{R\} \cup \{\omega_\ell\} \times \{NR\} \cup \overline{\{\omega_k, \omega_\ell\}} \times \mathcal{R}, \end{aligned} \quad (\text{A.15})$$

l'expression de $m_r^{\Omega\times\mathcal{R}}$ peut être obtenue :

$$m_r^{\Omega\times\mathcal{R}}(\overline{C} \times \{R\} \cup C \times \{NR\}) = \prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell, \quad \forall C \subseteq \Omega. \quad (\text{A.16})$$

Comme précédemment (A.6), la fonction de masse ci-dessus doit maintenant être combinée avec $m_{Ag}^{\Omega}[R]^{\uparrow\Omega\times\mathcal{R}}$, dont les éléments focaux sont de la forme $B \times \{R\} \cup \Omega \times \{NR\}$, avec $B \subseteq \Omega$. L'intersection des éléments focaux de $m_{Ag}^{\Omega}[R]^{\uparrow\Omega\times\mathcal{R}}$ et $m_r^{\Omega\times\mathcal{R}}$ est :

$$(\overline{C} \times \{R\} \cup C \times \{NR\}) \cap (B \times \{R\} \cup \Omega \times \{NR\}) = (B \cap \overline{C}) \times \{R\} \cup C \times \{NR\}.$$

Ainsi :

$$m_{Ag}[R]^{\uparrow\Omega\times\mathcal{R}} \odot m_r^{\Omega\times\mathcal{R}}((B \cap \overline{C}) \times \{R\} \cup C \times \{NR\}) = \left[\prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell \right] m_S^\Omega(B). \quad (\text{A.17})$$

Après projection sur Ω , on a :

$$\alpha m^\Omega(A) = \sum_{(B \cap \overline{C}) \cup C = A} \left[\prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell \right] m_S^\Omega(B), \quad \forall A \subseteq \Omega, \quad (\text{A.18})$$

$$= \sum_{B \cup C = A} \left[\prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell \right] m_S^\Omega(B), \quad \forall A \subseteq \Omega, \quad (\text{A.19})$$

et, l'équation A.8 de l'affaiblissement contextuel est retrouvée. \square

Cas 2 : Si l'équation 4.22 est remplacée par

$$\begin{cases} m_{Ag}^{\mathcal{R}}[\omega_k](\mathcal{R}) & = \beta_k, \\ m_{Ag}^{\mathcal{R}}[\omega_k](\{NR\}) & = \alpha_k. \end{cases} \quad (\text{A.20})$$

Dans ce cas, le déconditionnement de $m_{Ag}^{\mathcal{R}}[\omega_k]$ est :

$$m_{Ag}^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(\Omega \times \mathcal{R}) = \beta_k, \quad (\text{A.21})$$

$$m_{Ag}^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}(\{\omega_k\} \times \{NR\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) = \alpha_k. \quad (\text{A.22})$$

Notons $m_r^{\Omega\times\mathcal{R}}$ la combinaison conjonctive des $m_{Ag}^{\mathcal{R}}[\omega_k]^{\uparrow\Omega\times\mathcal{R}}$, $k = 1, \dots, K$. À partir de l'égalité suivante, $k \neq \ell$:

$$\begin{aligned} (\{\omega_k\} \times \{NR\} \cup \overline{\{\omega_k\}} \times \mathcal{R}) \cap (\{\omega_\ell\} \times \{NR\} \cup \overline{\{\omega_\ell\}} \times \mathcal{R}) = \\ \{\omega_k, \omega_\ell\} \times \{NR\} \cup \overline{\{\omega_k, \omega_\ell\}} \times \mathcal{R}. \end{aligned} \quad (\text{A.23})$$

l'expression de $m_r^{\Omega\times\mathcal{R}}$ peut être obtenue :

$$m_r^{\Omega\times\mathcal{R}}(C \times \{NR\} \cup \overline{C} \times \mathcal{R}) = \prod_{\omega_k \in C} \alpha_k \prod_{\omega_\ell \in \overline{C}} \beta_\ell. \quad (\text{A.24})$$

Cette fonction de masse doit être combinée avec $m_{Ag}^{\Omega}[R]^{\uparrow\Omega\times\mathcal{R}}$, dont les éléments focaux sont de la forme $B \times \{R\} \cup \Omega \times \{NR\}$, avec $B \subseteq \Omega$. L'intersection des éléments focaux de $m_{Ag}^{\Omega}[R]^{\uparrow\Omega\times\mathcal{R}}$ et $m_r^{\Omega\times\mathcal{R}}$ est :

$$(C \times \{NR\} \cup \overline{C} \times \mathcal{R}) \cap (B \times \{R\} \cup \Omega \times \{NR\}) = \Omega \times \{NR\} \cup (\overline{C} \cap B) \times \{R\}.$$

Ainsi, tous les éléments focaux de $m_{Ag}^{\Omega}[R]^{\uparrow\Omega\times\mathcal{R}} \odot m_r^{\Omega\times\mathcal{R}}$ contiennent $\Omega \times \{NR\}$. Par conséquent, en projetant sur Ω , la masse de chaque élément focal sera transféré à Ω , et

$$\alpha m_{Ag}^{\Omega}(\Omega) = (m_{Ag}^{\Omega}[R]^{\uparrow\Omega\times\mathcal{R}} \odot m_r^{\Omega\times\mathcal{R}})^{\downarrow\Omega}(\Omega) = 1. \quad (\text{A.25})$$

\square

A.3 Preuves des propositions 4.5 et 4.6

Notons $m_r^{\Omega\times\mathcal{R}}$ la combinaison conjonctive des $m_{Ag}^{\mathcal{R}}[\theta_k]^{\uparrow\Omega\times\mathcal{R}}$, $k = 1, \dots, L$. Un raisonnement similaire à la preuve de la proposition 4.1 montre que les éléments focaux de $m_r^{\Omega\times\mathcal{R}}$ sont de la forme $\overline{C} \times \{R\} \cup C \times \mathcal{R}$ avec $C \in \mathcal{C}$, et

$$m_r^{\Omega\times\mathcal{R}}(\overline{C} \times \{R\} \cup C \times \mathcal{R}) = \prod_{\cup\theta_k=C} \alpha_k \prod_{\cup\theta_\ell=\overline{C}} \beta_\ell, \quad (\text{A.26})$$

équation équivalente à (A.5).

Après la combinaison avec $m_{Ag}^\Omega[R]^{\uparrow\Omega \times \mathcal{R}}$ et marginalisation sur Ω , nous obtenons pour tout $A \subseteq \Omega$:

$$\begin{aligned} \mathfrak{G}m^\Omega(A) &= \sum_{B \cup C = A} m_r^{\Omega \times \mathcal{R}}(\overline{C} \times \{R\} \cup C \times \mathcal{R}) m_S^\Omega(B) \\ &= \sum_{B \cup C = A} \left[\prod_{\cup \theta_k = C} \alpha_k \prod_{\cup \theta_\ell = \overline{C}} \beta_\ell \right] m_S^\Omega(B) \\ &= \sum_{B \subseteq A} \mathfrak{G}G(A, B) m_S^\Omega(B), \end{aligned} \quad (\text{A.27})$$

où $\mathfrak{G}G(A, B)$ dénote un coefficient de la matrice de généralisation associée à l'affaiblissement Θ -contextuel :

$$\mathfrak{G}G(A, B) = \begin{cases} \sum_{B \cup C = A} \prod_{\cup \theta_k = C} \alpha_k \prod_{\cup \theta_\ell = \overline{C}} \beta_\ell & \text{si } \exists C \in \mathcal{C}, B \cup C = A, \\ 0 & \text{sinon.} \end{cases} \quad (\text{A.28})$$

Soient $A, B \subseteq \Omega$, et $C \in \mathcal{C}$ tels que $B \cup C = A$. Pour tout $\theta \subseteq C$, $\theta \subseteq B$ ou bien $\theta \cap (A \setminus B) \neq \emptyset$. Donc $C = (A \setminus B)^* \cup D$, où $D \in \mathcal{C}$, $D \subseteq B_*$, $C \in \mathcal{C}$ et $B \cup C = A$.

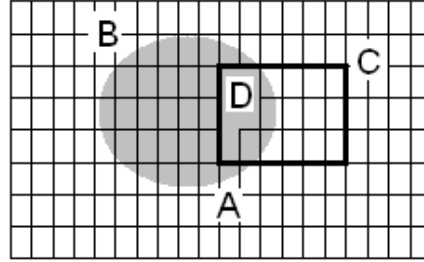


Figure A.1 – $C = (A \setminus B)^* \cup D$, où $D \in \mathcal{C}$, $D \subseteq B_*$, $C \in \mathcal{C}$ et $B \cup C = A$.

De même, pour tout $\theta \subseteq \overline{C}$, $\theta \subseteq (B \setminus D)$ ou bien $\theta \cap \overline{A} \neq \emptyset$. Ainsi, $\overline{C} = (B \setminus D)_* \cup \overline{A}^*$. Par conséquent :

$$\begin{aligned} \mathfrak{G}G(A, B) &= \sum_{D \subseteq B_*} \prod_{\cup \theta_k = (A \setminus B)^* \cup D} \alpha_k \prod_{\cup \theta_\ell = (B \setminus D)_* \cup \overline{A}^*} \beta_\ell \\ &= \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup \theta_\ell = \overline{A}^*} \beta_\ell \sum_{D \subseteq B_*} \prod_{\cup \theta_k = D} \alpha_k \prod_{\cup \theta_\ell = (B \setminus D)_*} \beta_\ell \\ &= \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup \theta_\ell = \overline{A}^*} \beta_\ell \prod_{\cup \theta_k = B_*} (\alpha_k + \beta_k) \\ &= \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup \theta_\ell = \overline{A}^*} \beta_\ell, \end{aligned}$$

ce qui conclut la preuve de la proposition 4.5.

La proposition 4.6 résulte de l'équation (A.27).

□

A.4 Preuve de la proposition 4.10

Avec les hypothèses réalisées, nous avons, $\forall i \in \llbracket 1, N \rrbracket$ et $\forall A \subseteq \Omega$:

$$m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}}(\Omega \times \{R_i\}) = \gamma_i, \quad (\text{A.29})$$

et

$$m_{Ag}^{\Omega}[R_i]^{\uparrow\Omega \times \mathcal{R}}(A \times \{R_i\} \cup \Omega \times \overline{\{R_i\}}) = m_i(A). \quad (\text{A.30})$$

De plus, $\forall i \in \llbracket 1, N \rrbracket$ et $\forall A_i \subseteq \Omega$:

$$\bigcap_{i=1}^N (A_i \times \{R_i\} \cup \Omega \times \overline{\{R_i\}}) = \bigcup_{i=1}^N A_i \times \{R_i\}, \quad (\text{A.31})$$

et, $\forall j \in \llbracket 1, N \rrbracket$:

$$\left(\bigcup_{i=1}^N A_i \times \{R_i\}\right) \cap \Omega \times \{R_j\} = A_j \times \{R_j\}. \quad (\text{A.32})$$

Ainsi, la combinaison conjonctive des $m_{Ag}^{\Omega}[R_i]^{\uparrow\Omega \times \mathcal{R}}$ avec $m_{Ag}^{\mathcal{R}\uparrow\Omega \times \mathcal{R}}$, que nous noterons $\odot m_{Ag}^{\Omega \times \mathcal{R}}$, possède N éléments focaux tels que :

$$\odot m_{Ag}^{\Omega \times \mathcal{R}}(A_j \times \{R_j\}) = \gamma_j m_j(A_j) \prod_{i \neq j} \underbrace{\sum_{A \subseteq \Omega} m_i(A)}_{=1}, \quad \forall j \in \llbracket 1, N \rrbracket, \quad (\text{A.33})$$

ou, de manière équivalente, $\forall A \subseteq \Omega$ et $\forall i \in \llbracket 1, N \rrbracket$:

$$\odot m_{Ag}^{\Omega \times \mathcal{R}}(A \times \{R_i\}) = \gamma_i m_i(A). \quad (\text{A.34})$$

Par conséquent, après projection sur Ω :

$$m_{Ag}^{\Omega}(A) = \sum_{i=1}^N \gamma_i m_i(A) \quad \forall A \subseteq \Omega. \quad (\text{A.35})$$

□

Titre :

Fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance

Résumé :

Dans ce mémoire, un problème de fusion de décisions postales est présenté. Une des caractéristiques de ce problème réside dans l'organisation hiérarchique des décisions, chaque lecteur d'adresses postales pouvant fournir une décision d'un niveau quelconque dans la hiérarchie. Plusieurs modèles de fusion de décisions postales sont construits dans le cadre de la théorie des fonctions de croyance. Des méthodes de construction d'une fonction de croyance à partir d'une matrice de confusion et d'une hiérarchie, sont exposées, ainsi que la méthode de prise de décision permettant un ajustement de la combinaison aux contraintes de taux de lecture et taux d'erreur attendues par l'utilisateur. Des performances probantes sont obtenues par ces modèles. Sur le plan théorique, de nouveaux outils de manipulation de fonctions de croyance ont été développés afin d'exploiter plus finement des informations sur la fiabilité des sources à combiner. Par l'affaiblissement contextuel, la dépendance au contexte de la fiabilité peut être pris en compte. Par le mécanisme général de correction proposé, différents états de fiabilité peuvent être modélisés, ce qui permet de développer des stratégies de correction autres qu'un affaiblissement.

Mots-clés :

Fusion d'adresses postales, Modèle des Croyances Transférables (MCT), Affectation Hiérarchique, Fiabilité de Capteur, Mécanisme de Correction d'une Fonction de Croyance.

Title :

Information fusion for the automatic recognition of postal addresses based on belief functions theory

Abstract:

In this report, a problem of postal addresses fusion is tackled. One characteristic of this problem is the hierarchical organisation of the decisions, each postal address reader being able to provide a decision at every level of the hierarchy. Several models are built within the framework of belief functions theory. Methods for building belief functions based on confusion matrices and the aforementioned hierarchy are exposed, as well as a decision-making method. This method allows an adjustment of the combination to the recognition and error rates awaited by the user. These models show convincing performances. On the theoretical level, new tools for the handling of belief functions are developed. In comparison with existing tools, the ones we propose improve the exploitation of information on the reliability of the sources to be combined. Contextual discounting allows the context dependence of the reliability to be taken into account. By the use of general correction mechanisms, various states of reliability can be modelled, which makes it possible to develop strategies of correction other than a weakening.

Keyword :

Postal addresses fusion, Transferable Belief Model (TBM), Hierarchical Assignment, Sensor Reliability, Belief Function Correction Mechanism.