

# Fusion de décisions postales dans le cadre du Modèle des Croyances Transférables

Postal decision fusion based on the Transferable Belief Model

**David Mercier<sup>1,2</sup>, Geneviève Cron<sup>2</sup>, Thierry Dencœur<sup>1</sup>, Marie-Hélène Masson<sup>1</sup>**

<sup>1</sup>UMR CNRS 6599 Heudiasyc, Université de Technologie de Compiègne, BP 20529 – F-60205 Compiègne cedex - France

<sup>2</sup>SOLYSTIC, 14 avenue Raspail, F-94257 Gentilly Cedex, France

Manuscrit reçu le 11 mai 2006

Résumé et mots clés

Afin d'améliorer la prise de décision pour la reconnaissance automatique d'adresses postales, un modèle de fusion de décisions de lecteurs d'adresses postales (LAP) basé sur la théorie des fonctions de croyance est exposé. Ce cadre théorique offre une grande flexibilité dans la représentation et la combinaison des décisions fournies par les LAP, chacun pouvant fournir une adresse postale complète ou partielle. Dans cet article, nous détaillons le cadre de discernement choisi, l'affectation des masses basée sur une organisation hiérarchique des décisions, et la méthode de prise de décision adoptée. Les résultats obtenus montrent une réelle amélioration par rapport aux performances individuelles des LAP.



Fusion d'adresses postales, Théorie de Dempster-Shafer, Modèle des Croyances Transférables (MCT), Hiérarchie de décisions.

Abstract and key words

In order to improve decision-making in automatic recognition of postal addresses, a model for the fusion of postal address recognition systems (or OCR engine, for optical character recognition) based on belief functions theory is exposed. This theoretical framework offers a great flexibility in the representation and combination of the decisions provided by the OCRs, each one being able to provide a complete or a partial postal address. In this article, we describe the chosen frame of discernment, the mass assignment developed from a hierarchical organization of the decisions and the decision-making method. Experimental results show a real improvement of the LAPs' individual performances.

Postal Addresses Fusion, Dempster-Shafer Theory, Transferable Belief Model (TBM), hierarchy of decisions.

# 1. Introduction

La reconnaissance automatique d'adresses postales est un problème difficile, recourant à une chaîne complexe de traitements variable suivant les pays [2]. Cette chaîne débute par l'acquisition d'une image d'un courrier sur lequel se trouve l'adresse du destinataire à identifier. Elle comporte 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)*.

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, 3, 10, 22].

Les solutions existantes abordent le problème à partir de procédures de votes à la majorité ou pondérés, ou d'heuristiques de combinaison [4, 9]. Une autre approche est proposée ici, basée sur le *modèle des croyances transférables (MCT)* [21, 17]. Le MCT est un modèle de représentation de connaissances partielles reposant sur une interprétation subjective de la théorie des fonctions de croyance de Dempster et Shafer [13, 16]. Il permet de gérer des informations aussi bien incertaines qu'imprécises, et fournit nombre d'outils pour la combinaison d'informations [18].

Dans cet article, nous décrivons la problématique de fusion d'adresses postales et montrons comment un modèle basé sur le MCT peut s'appliquer. Un accent particulier est mis sur la méthode d'affectation des masses réalisée à partir des matrices de confusion des LAP. La méthode présentée s'inspire de celle de Xu, Krzyzak et Suen [23] en l'étendant à un problème où les informations à combiner sont organisées de manière hiérarchique.

L'organisation de cet article est la suivante. Dans le paragraphe 2, les concepts de base du MCT sont rappelés. Le paragraphe 3 aborde la problématique des informations postales à fusionner. Le paragraphe 4 décrit le choix du cadre de discernement. Les paragraphes 5 et 6 présentent les trois grandes étapes du modèle de combinaison: la génération des fonctions de croyance, la combinaison des fonctions de croyance, et la prise de décision permettant d'obtenir des décisions à plusieurs niveaux. Le paragraphe 7 présente les résultats expérimentaux obtenus par la combinaison de trois LAP par cette méthode de combinaison. Enfin, le paragraphe 8 conclut et énonce les perspectives.

# 2. Le Modèle des Croyances Transférables

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 :

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

## 2.1. Représentation des connaissances

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. En général, les masses de croyance sont cependant allouées à des sous-ensembles plus grands que des singletons, et, par manque d'informations, ne peuvent pas être affectées à des sous-ensembles plus restreints.

Formellement, soit  $x$  une variable à valeurs dans un ensemble fini  $\Omega = \{\omega_1, \dots, \omega_K\}$ , appelé *cadre de discernement* ou *univers*. La connaissance détenue par un agent, quant à la valeur  $\omega_0$  effectivement prise par  $x$ , peut être quantifiée par une *fonction de masse* définie sur l'ensemble des parties de  $\Omega$ , à valeurs dans  $[0, 1]$ , et vérifiant  $\sum_{A \subseteq \Omega} m(A) = 1$ .

La quantité  $m(A)$  représente la *part de croyance* de l'agent sur le fait « $\omega_0 \in A$ ».

Notons que dans le MCT, une fonction de masse peut être non normalisée: il est possible que  $m(\emptyset)$  soit strictement positive. L'interprétation de  $m(\emptyset)$  est discutée dans [14, 20].

La *fonction de masse vide* est définie par  $m^\Omega(\Omega) = 1$ . Elle représente l'ignorance totale sur la valeur de  $x$ . Un sous-ensemble  $A$  de masse non nulle est appelé *élément focal*. Une *fonction de masse Bayésienne* est une fonction de masse dont les éléments focaux sont des singletons. Une *fonction de masse consonante* est une fonction de masse dont les éléments focaux sont emboîtés.

## 2.2. Manipulation des connaissances

Considérons une masse  $m(A)$  strictement positive allouée à un sous-ensemble  $A$  de  $\Omega$ . Lorsque nous apprenons qu'avec certitude la vérité se situe dans un sous-ensemble  $B$  de  $\Omega$ , 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 «*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, les fonctions de croyance peuvent être déconditionnées, marginalisées, étendues, affaiblies ou combinées entre elles [15, 21]. Ce large spectre d'outils de manipulation des fonctions de croyance explique l'intérêt de cette théorie pour la fusion d'informations.

Le modèle présenté dans cet article exploite principalement le pouvoir de représentation des fonctions de croyance. Leur manipulation au niveau crédal se limitera à la *règle de combinaison conjonctive*.

Deux fonctions de masse,  $m_1$  et  $m_2$ , issues de sources d'information distinctes et fiables peuvent être combinées en utilisant la *règle de combinaison conjonctive* définie par :

$$m_1 \odot m_2(A) = \sum_{B \cap C = A} m_1(B)m_2(C), \quad \forall A \subseteq \Omega. \quad (1)$$

### 2.3. Niveau pignistique

Lorsqu'une décision doit être prise, des *principes de rationalité* [5, 12] justifient la stratégie consistant à choisir la décision  $d$  parmi un ensemble de décisions possibles  $\mathcal{D}$  minimisant le *risque espéré* défini par :

$$\rho(d) = \sum_{\omega \in \Omega} c(d, \omega) P^\Omega(\{\omega\}), \quad (2)$$

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. La quantité  $c(d, \omega)$  représente le coût de décider  $d$  sachant que la vérité est  $\omega$ .

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* [6, 8, 21, 19] définie par :

$$\text{Bet } P(\{\omega\}) = \sum_{\{A \subseteq \Omega, \omega \in A\}} \frac{m(A)}{|A| (1 - m(\emptyset))}. \quad (3)$$

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 [19]. La solution obtenue consiste à répartir uniformément la masse de chaque élément focal entre les singletons qui le composent. En présence 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, cette stratégie conduit à choisir l'action de probabilité pignistique maximale.

Au niveau pignistique, lorsqu'une décision doit être prise, une fonction de probabilité est construite à partir de la fonction de croyance combinant toutes les informations disponibles au niveau crédal.

## 3. Problématique de la fusion de décisions postales

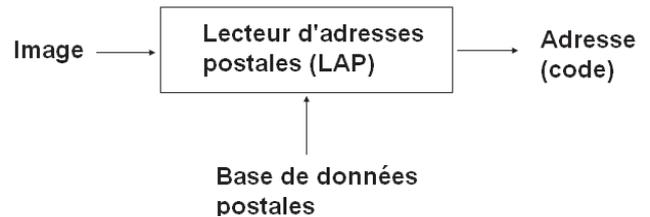


Figure 1. Modélisation d'un lecteur d'adresses postales.

Un *lecteur d'adresses postales (LAP)* peut être vu comme un processus lié à une base de données d'adresses postales, prenant en entrée une image d'un envoi postal (lettre, courrier, colis, etc.) sur laquelle se trouve l'adresse postale à identifier, et retournant un code, appelé *code d'indexation* (figure 1).

Une adresse postale est composée d'une *partie acheminement* et d'une *partie distribution*. La partie acheminement indique le bureau de distribution, en général la ville de destination de l'envoi postal. La partie distribution indique le lieu précis de délivrance de l'envoi. Elle est relative, soit à un pas de porte dans une rue et dans ce cas nous parlerons de *distribution géographique*, soit à des *distributions spéciales* assimilables à des boîtes postales.

Dans cet article, nous utiliserons des notations simplifiées pour le code d'indexation. Nous noterons une partie acheminement par  $A_i$ , une boîte postale ou tout autre distribution spéciale par  $B_j$ , une rue ou tout autre voie géographique (avenue, boulevard, place, etc.) par  $R_k$ , où  $i, j$  et  $k$  sont des indices. Le numéro de pas de porte dans une rue se notera à la fin du code d'indexation. Ainsi, « $A_2R_26$ » codera 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, « $A_3B_1$ » codera la boîte postale  $B_1$  dans  $A_3$ .

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 « $A_1$ » pourrait être associé à l'adresse «2 le Bourg – 33230 Saint Christophe de Double». Ces lieux d'acheminement seront dit *non distribuables* : il n'est pas nécessaire de travailler sur la distribution pour obtenir le code de l'adresse complète.

En distribution géographique, il est possible que tous les pas de porte d'une voie soient codés par le même code distribution. Dans ce cas, le numéro de pas de porte n'apparaît pas dans le code d'indexation : la ville et la rue suffisent à obtenir le code de l'adresse complète. Par exemple, une sortie de LAP en faveur du code « $A_3R_1$ » codera tous les pas de porte de la rue  $R_1$  dans  $A_3$ . Ces voies seront dites *non numérotées*.

Pour résumer, les sorties suivantes de LAP sont associées à des adresses complètes :

- «  $A_1$  » : un code acheminement non distribuable ;
- «  $A_3 R_1$  » : une voie non numérotée ;
- «  $A_2 R_2 2$  » : une voie numérotée ;
- «  $A_3 B_1$  » : un point de distribution spéciale.

La reconnaissance complète d'une adresse postale étant une tâche complexe, un LAP est autorisé à fournir des adresses partielles. Par exemple, un LAP relativement sûr de la reconnaissance de la ville (par exemple  $A_2$ ), mais dont la reconnaissance de la rue pose problème pourra proposer la ville et « rejeter la rue ». Cette décision sera notée par le code «  $A_2 Rej$  ». La notation « *Rej* » indiquera le rejet sur une partie devant être précisée pour obtenir un code associé à une adresse complète. Un code comportant un rejet sera donc associé à une adresse partielle.

Ainsi, les sorties suivantes de LAP sont associées à des adresses partielles :

- « *Rej* » : un rejet complet ;
- «  $A_3 Rej$  » : un rejet distribution ;
- «  $A_2 R_2 Rej$  » : un rejet d'un numéro de pas de porte.

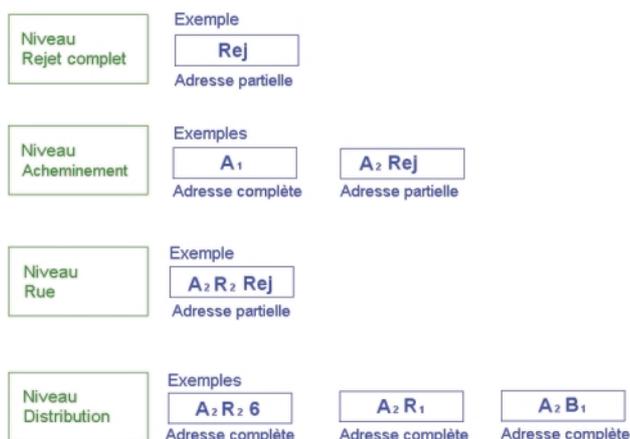


Figure 2. Niveaux des décisions autorisées pour chaque LAP avec des exemples associés.

Le niveau sur lequel peut s'arrêter la décision d'un LAP se fait en accord avec une hiérarchie de décisions, dont la définition formelle sera donnée un peu plus loin au paragraphe 5.2. Les niveaux de la hiérarchie utilisée en exemple dans cet article sont illustrés sur la figure 2.

Supposons disposer de  $N$  LAP  $LAP_i, i \in \{1, \dots, N\}$  associés à la même hiérarchie. La problématique de fusion consiste à fournir, pour chaque image, la « meilleure décision » possible connaissant les  $N$  décisions issues des  $N$  LAP. Pour statuer sur cette meilleure décision, des critères de performance, comme le *taux de lecture* (taux de bonnes décisions), le *taux d'erreur* ou le *taux de rejet*, sont utilisés à chaque niveau de la hiérarchie. Généralement, les LAP sont comparés selon leurs taux de performance (taux de lecture, taux 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 «  $A_2 R_1$  » fournie pour une image dont la vérité est «  $A_2 R_2 1$  » est comptée comme une bonne réponse en acheminement : en effet, la lettre sera acheminée dans la bonne ville. Mais, cette décision sera comptée comme fautive en 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. Par exemple, sur le courrier dactylographié, les LAP à combiner ont individuellement un taux de lecture en acheminement supérieur à 95 % pour un taux d'erreur inférieur à 0.5 %, et un taux de lecture en distribution supérieur à 85 % pour un taux d'erreur inférieur à 1 %.

L'objectif des modèles de combinaison réside dans l'obtention de performances plus élevées que celles des LAP à fusionner. Les trois grandes étapes d'un tel modèle de combinaison :

1. la génération des fonctions de croyance ;
2. la combinaison des fonctions de croyance ;
3. la prise de décision,

sont décrites dans les paragraphes 5 et 6.

Dans le paragraphe suivant, le choix du cadre de discernement, une étape importante de la genèse de tout modèle utilisant des fonctions de croyance, est décrit.

## 4. 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. Cependant, 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 :

- 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.

Si l'adresse inscrite sur l'enveloppe ne contient pas d'erreurs, alors cette adresse sera dite *valide*. Les adresses comportant des erreurs 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.

Avec les niveaux considérés sur la figure 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.

Ainsi, le cadre de discernement  $\Omega$ , 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  $\Omega$  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 «  $A_2inv$  ».

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 forte-

ment que la vérité est «  $A_2R_2inv$  », le numéro de pas de porte étant invalide, il va décider «  $A_2R_2Rej$  ». 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 4.1 (Sur une base de données réduite)

Considérons une base de données ne contenant que deux points d'acheminement,  $A_1$  et  $A_2$ , où  $A_1$  est non distribuable, et  $A_2$  est distribuable.  $A_2$  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  $\Omega$  peuvent être représentés sur la figure 3. Les singletons de  $\Omega$  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  $A_2R_2 = \{A_2R_22, A_2R_24, A_2R_26, A_2R_2inv\}$ .

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  $\Omega$  :

- $\{A_1\} = \langle A_1 \rangle$ ,
- $\{A_2B_1\} = \langle A_2B_1 \rangle$ ,
- $\{A_2B_2\} = \langle A_2B_2 \rangle$ ,
- $\{A_2R_1\} = \langle A_2R_1 \rangle$ ,
- $\{A_2R_22\} = \langle A_2R_22 \rangle$ ,
- $\{A_2R_24\} = \langle A_2R_24 \rangle$ ,
- $\{A_2R_26\} = \langle A_2R_26 \rangle$ , et,

- des décisions en faveur d'adresses partielles, qui sont des sous-ensembles de  $\Omega$  :

- $\Omega = \langle Rej \rangle$ ,
- $A_2 = \langle A_2Rej \rangle$ ,
- $A_2R_2 = \langle A_2R_2Rej \rangle$ .

Toute information sur la véritable adresse sera modélisée par une fonction de croyance sur le cadre de discernement  $\Omega$ . 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  $\Omega$ . 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.

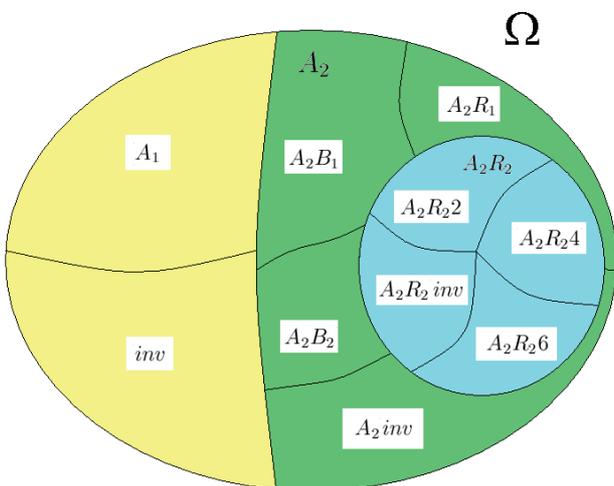


Figure 3. Éléments du cadre de discernement associés à l'exemple.

# 5. Génération des fonctions de croyance

Disposant des décisions des LAP, il est possible d'étudier des pourcentages de réussites associés à ces décisions à partir d'un ensemble d'apprentissage composé d'images dont la vérité est connue. Une matrice de confusion attachée à un classifieur regroupe tous ces résultats passés. Deux principales méthodes d'affectation basées sur une décision et une matrice de confusion sont présentées dans le paragraphe suivant.

## 5.1. Matrices de confusion et affectations classiques

Formellement, une *matrice de confusion*  $M = (n_{k\ell})_{k \in \{1, \dots, K+1\} \ell \in \{1, \dots, K\}}$  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  $\omega_k$ . Chaque colonne  $\ell$  correspond au cas où la vérité est  $\omega_\ell$ . Le terme général  $n_{k\ell}$  est égal au nombre d'objets de classe  $\omega_\ell$  ayant été classés dans la classe  $\omega_k$  par le classifieur.

La décision «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, pour toute la suite de cet article, un rejet complet est identifié à une décision en faveur de l'univers  $\Omega$  tout entier. Uniquement dans ce paragraphe, dans un but de simplicité d'écriture des expressions, une décision en faveur d'un rejet est notée comme une décision en faveur d'une classe spéciale  $\omega_{K+1}$ . Ainsi la matrice de confusion est de dimension  $(K + 1, K)$  (Tableau 1).

Tableau 1. Illustration d'une matrice de confusion.

Décision	Vérité	$\omega_1$	...	$\omega_K$
	$\omega_1$		$n_{11}$	...
$\vdots$		$\vdots$	$\ddots$	$\vdots$
$\omega_K$		$n_{K1}$	...	$n_{KK}$
$\omega_{K+1}$		$n_{(K+1)1}$	...	$n_{(K+1)K}$

Notons  $n_{k\cdot} = \sum_{\ell=1}^K n_{k\ell}$  le nombre d'objets classés dans  $\omega_k$ , 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  $\Omega$  ;

- l'affectation de Xu *et al.* [23] allouant de la masse à la classe décidée par le classifieur, à son complémentaire et à l'ignorance.

### 5.1.1. 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  $\omega_\ell$  ayant été classés dans  $\omega_k$ , divisé par le nombre total d'objets classés dans  $\omega_k$ :

$$\forall \omega_\ell \in \Omega, m(\{\omega_\ell\}) = \frac{n_{k\ell}}{\sum_{j=1}^K n_{kj}} = \frac{n_{k\ell}}{n_{k\cdot}}. \tag{4}$$

#### Exemple 5.1

Considérons un classifieur en charge de la reconnaissance des caractères numériques, dont la matrice de confusion est illustrée dans le tableau 2. 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  $m(\{1\}) = 1/20, m(\{2\}) = 2/20, m(\{3\}) = 3/20, m(\{4\}) = 4/20, m(\{5\}) = 5/20, m(\{6\}) = 2/20, m(\{7\}) = 3/20$ .

Tableau 2. 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
D	1	10	0	0	0	0	0	0	0	0
	2	0	8	0	0	0	0	0	0	0
	E	3	0	0	3	0	0	0	0	0
	C	4	0	0	0	4	0	0	0	0
	I	5	0	0	0	0	5	0	0	0
	S	6	0	0	0	0	0	6	0	2
	I	7	0	4	0	0	0	0	7	0
	O	8	0	0	0	0	0	0	0	8
	N	9	0	0	2	4	0	0	0	0
	Rejet	1	2	3	4	5	2	3	0	

Une telle affectation conduit à des fonctions de masse Bayésiennes. En conséquence, pour que cette précision soit justifiée et que le rapport  $n_{k\ell}/n_{k\cdot}$  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.

5.1.2. Affectation de Xu *et al.* [23]

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

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

$$R = \frac{\sum_{k=1}^K n_{kk}}{n}, \tag{5}$$

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

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

- le *taux de rejet*  $T$  est égal à

$$T = \frac{n_{(K+1)}}{n}, \tag{7}$$

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

Quand un classifieur fournit une décision  $d = \{\omega_k\}$  avec  $k \in \{1, \dots, K\}$ , il est proposé dans [23] d'utiliser ces trois taux pour définir  $m$  de la façon suivante :

$$m : \begin{array}{ll} 2^\Omega & \longrightarrow [0,1] \\ \{\omega_k\} & \longmapsto R \\ \Omega \setminus \{\omega_k\} & \longmapsto S \\ \Omega & \longmapsto T \end{array} \tag{8}$$

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

**Exemple 5.2**

En reprenant le classifieur associé à la matrice de confusion illustrée sur le tableau 2. 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 [23] 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 5.3**

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.* (8) 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  $C_1$ , *i.e.*, 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.*

5.1.3. Une affectation basée sur le taux de fiabilité

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

$$F = \frac{\sum_{k=1}^K n_{kk}}{\sum_{k=1}^K \sum_{\ell=1}^K n_{k\ell}} = \frac{R}{1 - T}. \tag{9}$$

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 :

$$m : \begin{array}{ll} 2^\Omega & \longrightarrow [0,1] \\ \{\omega_k\} & \longmapsto F \\ \Omega & \longmapsto 1 - F \end{array} \tag{10}$$

De même que l'affectation de Xu *et al.* (8), quand un rejet est décidé,  $m(\Omega) = 1$ .

L'affectation basée sur le taux de fiabilité (10) prend uniquement en compte la fiabilité des classifieurs quand une classe est choisie.

**Exemple 5.4**

Appliquée à l'exemple 5.3, 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, présentées dans le paragraphe suivant.

**5.2. Modèles hiérarchiques d'affectation**

Dans le paragraphe précédent, les classifieurs pouvaient décider une classe de  $\Omega$  ou un rejet. Dans ce paragraphe, les classifieurs

ont la possibilité de décider un ensemble de classes intermédiaire entre  $\Omega$  et un singleton. Par exemple, dans notre application postale, les LAP peuvent décider une adresse complète composée d'un numéro de pas de porte, d'une rue et d'une ville, mais par manque d'informations suffisantes sur le numéro de pas de porte, ce LAP peut décider la rue et la ville en ne précisant pas le numéro de pas de porte; de même, en absence d'informations sur la rue, il peut fournir une adresse partielle en précisant uniquement la ville du destinataire.

Dans le paragraphe suivant nous présentons la forme purement hiérarchique de l'affectation développée dans ce cadre de décisions multi-niveaux. Dans un deuxième temps, nous présentons une extension de ce modèle d'affectation où à chaque niveau de la hiérarchie des ensembles de classes peuvent être regroupés en catégorie.

### 5.2.1. Affectation hiérarchique

Une *hiérarchie* est un arbre dont la racine est  $\Omega$ , tel que l'ensemble des éléments de chaque *niveau* forme une partition de  $\Omega$ . Par convention, on considère que le premier niveau est le niveau composé des singletons de  $\Omega$  et le dernier niveau est celui composé de  $\Omega$ .

Nous noterons :

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

La figure 4 illustre un exemple de hiérarchie à trois niveaux, et la hiérarchie de l'exemple 4.1 lié à une base de données réduite est représentée sur la figure 5.

Dans ce contexte de décisions multi-niveaux, il est possible de distinguer plusieurs types d'erreurs. Par exemple, supposons qu'un LAP fournisse une décision  $\{A_2R_26\}$  située au niveau distribution de la hiérarchie représentée sur la figure 5. 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 vérité se trouve dans le même élément  $A_2R_2$  situé juste au dessus dans la hiérarchie. Ainsi, la ville et la rue de la décision fournie sont communes avec la véritable adresse. Nous dirons que cette décision, incorrecte au niveau distribution, est correcte au niveau rue. 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  $\Omega$  : cette décision, incor-

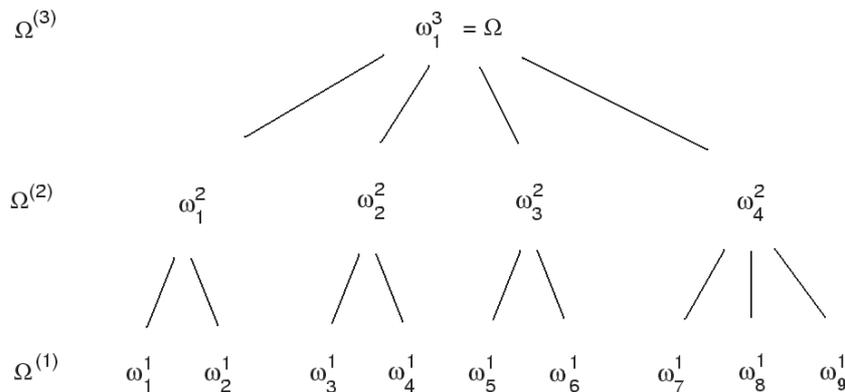


Figure 4. Exemple d'une hiérarchie à trois niveaux.

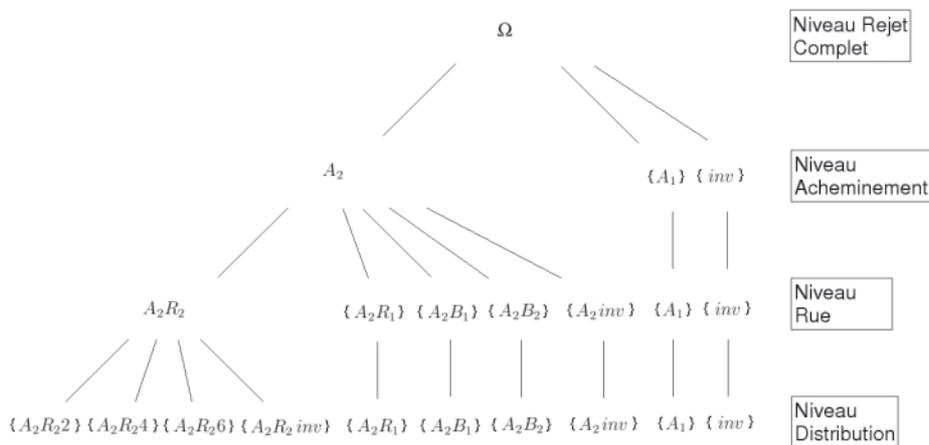


Figure 5. Hiérarchie associée à l'exemple 4.1 lié à une base de données réduite.

Tableau 3. Un exemple de matrice de confusion associée à un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 4.

Vérité	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	$\omega_7$	$\omega_8$	$\omega_9$
<b>Décisions</b>									
<b>Niveau 1</b>									
$\omega_1^1 = \{\omega_1\}$	<b>34</b>	<u>1</u>	1	0	0	0	0	0	0
$\omega_2^1 = \{\omega_2\}$	<u>2</u>	<b>23</b>	0	0	0	0	1	0	0
$\omega_3^1 = \{\omega_3\}$	0	1	<b>24</b>	<u>2</u>	0	0	0	0	0
$\omega_4^1 = \{\omega_4\}$	1	0	<u>3</u>	<b>28</b>	0	0	0	0	0
$\omega_5^1 = \{\omega_5\}$	0	0	0	1	<b>35</b>	<u>0</u>	0	1	0
$\omega_6^1 = \{\omega_6\}$	0	0	0	1	<u>3</u>	<b>15</b>	0	0	1
$\omega_7^1 = \{\omega_7\}$	0	2	3	0	4	1	<b>20</b>	<u>10</u>	<u>1</u>
$\omega_8^1 = \{\omega_8\}$	0	1	0	0	3	2	<u>5</u>	<b>35</b>	<u>1</u>
$\omega_9^1 = \{\omega_9\}$	0	1	0	3	2	0	<u>8</u>	<u>3</u>	<b>31</b>
<b>Niveau 2</b>									
$\omega_1^2 = \{\omega_1, \omega_2\}$	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
<b>Niveau 3</b>									
$\omega_1^3 = \Omega$	1	2	0	4	1	2	1	0	2

recte 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 est l'ensemble de classes  $A_2$ , donc en faveur d'une adresse partielle «  $A_2Rej$  », et la vérité est  $A_2R_2$ , cette décision de niveau acheminement est correcte à ce niveau. Pour chaque décision, il est possible de déterminer à quel niveau de la hiérarchie celle-ci est correcte. 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 5.5

Considérons un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 4 et dont la matrice de confusion est donnée dans le tableau 3. 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  $\Omega$  et les singletons.

Le nombre de décisions de niveau 1 correctes à ce niveau 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 égale 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  $\Omega$  pour être dans le même élément que la

vérité. L'élément parent de la décision ne contient pas la vérité. Ainsi :

$$\begin{aligned}
 n^{1,1} &= 34 + 23 + 24 + 28 + 35 + 15 + 20 + 35 + 31 \\
 &= 245, \\
 n^{1,2} &= 1 + 2 + 2 + 3 + 0 + 3 + 10 + 1 + 5 + 1 + 8 + 3 \\
 &= 39, \\
 n^{1,3} &= 1 \times 8 + 2 + 3 + 4 + 1 + 1 + 3 + 2 + 1 + 3 + 2 \\
 &= 30.
 \end{aligned} \tag{11}$$

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 :

$$n^p = \sum_{q=p}^P n^{p,q}. \tag{12}$$

Soit  $u$  la fonction qui à un élément de la hiérarchie différent de  $\Omega$ , associe son élément parent situé juste au dessus dans la hiérarchie. Par exemple, avec la hiérarchie représentée sur la figure 4,  $u(\omega_3^1) = \omega_2^2$ , 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 :

$$\begin{aligned}
 m : \quad 2^\Omega &\longrightarrow [0, 1] \\
 u^{q-p}(d) &\longmapsto \frac{n^{p,q}}{n^p} \quad \forall q \in [p, P],
 \end{aligned} \tag{13}$$

où  $u^{q-p} = \underbrace{u \circ \dots \circ u}_{q-p \text{ fois}}$ , 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 5.6 (suite de l'exemple 5.5)**

Reprenons le classifieur de l'exemple 5.5, et supposons qu'il décide  $\omega_3^1 = \{\omega_3\}$ . Cette décision est de niveau 1. Le nombre total de décisions de niveau 1 fournies par la classifieur est  $n^1 = n^{1,1} + n^{1,2} + n^{1,3} = 314$ . L'affectation (13) conduit à représenter la décision de ce classifieur par la fonction de masse  $m$ , telle que :

$$\begin{aligned}
 m(\omega_3^1) &= m(\{\omega_3\}) &= n^{1,1}/n^1 \\
 &= 245/314 &= 0.78, \\
 m(u(\omega_3^1)) &= m(\{\omega_3, \omega_4\}) &= n^{1,2}/n^1 \\
 &= 39/314 &= 0.12, \\
 m(u^2(\omega_3^1)) &= m(\Omega) &= n^{1,3}/n^1 \\
 &= 30/314 &= 0.10.
 \end{aligned}
 \tag{14}$$

5.2.2. 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 5, 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 étendue 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. Chaque décision des classifieurs sera un élément situé à un niveau de la hiérarchie, et appartiendra à une certaine catégorie. La hiérarchie représentée sur la figure 6 illustre une hiérarchie à trois niveaux avec des exemples de catégories de décisions.

L'affectation hiérarchique catégorisée se base sur la même idée de distance que l'affectation précédente. Les mêmes éléments

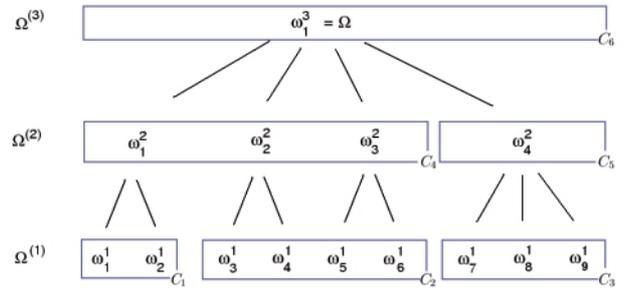


Figure 6. Un exemple de hiérarchie à trois niveaux avec des catégories de décisions.

focaux seront créés. Seules les valeurs des masses seront dépendantes du niveau de la décision fournie par le classifieur et de la catégorie de décision.

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

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

$$n_c^p = \sum_{q=p}^P n_c^{p,q}. \tag{15}$$

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

$$m : \begin{aligned}
 2^\Omega &\longrightarrow [0, 1] \\
 u^{q-p}(d) &\longmapsto \frac{n_c^{p,q}}{n_c^p} \quad \forall q \in [p, P].
 \end{aligned} \tag{16}$$

Notons que si  $d = \omega_1^p = \Omega$ , comme une seule catégorie est toujours définie au dernier niveau de la hiérarchie :  $m(d) = m(\Omega) = \frac{n_c^{p,p}}{n_c^p} = \frac{n^{p,p}}{n^p} = \frac{n^p}{n^p} = 1$ .

Les fonctions de masse générées par cette affectation demeurent consonantes.

**Exemple 5.7**

Considérons un classifieur fournissant des décisions en accord avec la hiérarchie de la figure 6 et dont la matrice de confusion, fournie dans le tableau 4, comporterait les mêmes valeurs que dans l'exemple 5.6 (tableau 3). 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  $c_3$  correctes au niveau 1 est égal à la somme des nombres inscrits en gras dans la matrice de la table 4. Le nombre de décisions de niveau 1 et de catégorie  $c_3$  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.

Tableau 4. Matrice de confusion d'un classifieur fournissant des décisions en accord avec la hiérarchie représentée sur la figure 6.

Vérité	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	$\omega_7$	$\omega_8$	$\omega_9$
<b>Décisions</b>									
<b>Niveau 1</b>									
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	<b>24</b>	<u>2</u>	0	0	0	0	0
$\omega_4^1 = \{\omega_4\}$	1	0	<u>3</u>	<b>28</b>	0	0	0	0	0
$\omega_5^1 = \{\omega_5\}$	0	0	0	1	<b>35</b>	<u>0</u>	0	1	0
$\omega_6^1 = \{\omega_6\}$	0	0	0	1	<u>3</u>	<b>15</b>	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	<b>31</b>
<b>Niveau 2</b>									
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
<b>Niveau 3</b>									
catégorie $c_6$									
$\omega_1^3 = \Omega$	1	2	0	4	1	2	1	0	2



$$\begin{aligned}
 n_{c_2}^{1,1} &= 24 + 28 + 35 + 15 = 102, \\
 n_{c_2}^{1,2} &= 2 + 3 + 0 + 3 = 8, \\
 n_{c_2}^{1,3} &= 1 + 1 + 1 + 1 + 1 + 1 = 6,
 \end{aligned}
 \tag{17}$$

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

$$\begin{aligned}
 m(\omega_3^1) &= m(\{\omega_3\}) = n_{c_2}^{1,1}/n_{c_2}^1 = 102/116 = 0.88, \\
 m(u(\omega_3^1)) &= m(\{\omega_3, \omega_4\}) = n_{c_2}^{1,2}/n_{c_2}^1 = 8/116 = 0.07, \\
 m(u^2(\omega_3^1)) &= m(\Omega) = n_{c_2}^{1,3}/n_{c_2}^1 = 6/116 = 0.05.
 \end{aligned}
 \tag{18}$$

Par rapport au résultat obtenu (14) avec l'affectation hiérarchique précédente, basée uniquement sur le niveau des décisions, la masse allouée à  $\{\omega_3\}$  est plus importante. La décision  $\omega_3^1 = \{\omega_3\}$  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 :

$$\begin{aligned}
 m(\omega_9^1) &= m(\{\omega_9\}) = n_{c_3}^{1,1}/n_{c_3}^1 = 86/136 = 0.63, \\
 m(u(\omega_9^1)) &= m(\{\omega_7, \omega_8, \omega_9\}) = n_{c_3}^{1,2}/n_{c_3}^1 = 28/136 = 0.21, \\
 m(u^2(\omega_9^1)) &= m(\Omega) = n_{c_3}^{1,3}/n_{c_3}^1 = 22/136 = 0.16.
 \end{aligned}
 \tag{19}$$

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

- Si la hiérarchie ne comporte qu'une seule catégorie par niveau, l'affectation catégorisée (16) est équivalente à l'affectation hiérarchique (13);
- Si la hiérarchie ne comporte que deux niveaux avec un premier niveau composé de tous les singletons de  $\Omega$  tous de même catégorie, alors l'affectation catégorisée (16) est équivalente à l'affectation basée sur le taux de fiabilité (10);
- Si la hiérarchie ne comporte que deux niveaux avec un premier niveau des singletons de  $\Omega$ , tous attachés à une catégorie propre, alors l'affectation développée (16) coïncide avec l'af-

fectation Bayésienne sur la décision du classifieur mais diffère sur les autres éléments. L'affectation (16) fournit toujours une masse consonante.

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

### 5.2.3. Affectation postale utilisée

La hiérarchie postale (figure 7) employée dans cet article, indique uniquement les niveaux et les catégories de décisions considérées, les éléments de  $2^\Omega$  ne peuvent pas être représentés du fait de leur trop grand nombre. Toutefois, sur l'exemple 4.1, relatif à une base de données réduite, il a été possible de représenter tous les éléments de  $\Omega$  dans la hiérarchie représentée sur la figure 5. Le lien avec la hiérarchie de référence (figure 7) est

illustré sur la figure 8. Nous pouvons remarquer que, dans cette hiérarchie postale, des éléments singletons peuvent être présents à plusieurs niveaux. Afin de déterminer de manière unique le niveau d'une décision en faveur d'un singleton pour l'affectation des masses, il a été considéré par convention qu'une décision en faveur d'un acheminement non distribuable était de niveau acheminement, et qu'une décision en faveur d'une boîte postale ou d'une voie non numérotée était de niveau distribution.

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

1. ajouter un niveau pays (un niveau intermédiaire entre le rejet complet et l'acheminement) afin de gérer les pays étrangers ;

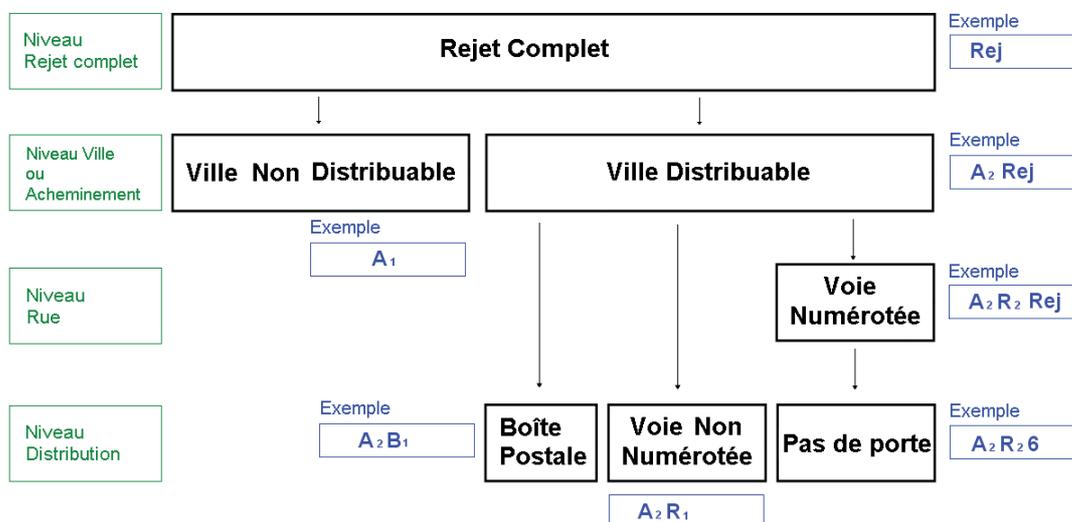


Figure 7. Hiérarchie et catégories des décisions postales avec des exemples de décision associés.

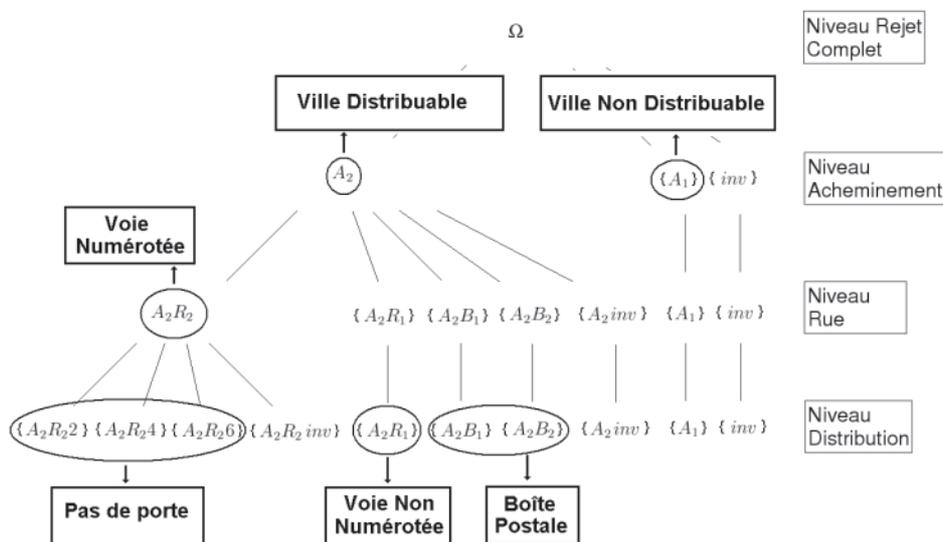


Figure 8. Catégories des décisions des LAP dans l'exemple 4.1.

Tableau 5. Matrice de confusion associée à un LAP sur la base de données réduite de l'exemple 4.1.

Niveau	Vérité	<i>inv</i>	$A_1$	$A_2$	$B_1$	$B_2$	$R_1$	$R_22$	$R_24$	$R_26$	$R_2$
	Décision	<i>inv</i>									
<sup>(1)</sup> $\Omega$ (Distri- bution)	$B_1$	0	0	1	95	3	0	0	0	1	0
	$B_2$	1	0	0	2	96	0	0	0	0	1
	$R_1$	0	1	0	0	0	97	1	1	0	0
	$R_22$	0	0	0	3	0	0	91	1	5	0
	$R_24$	0	2	0	0	0	3	6	88	1	0
	$R_26$	0	4	0	0	1	2	7	1	85	0
<sup>(2)</sup> $\Omega$ (Rue)	$R_2$	0	1	0	0	0	1	17	31	42	8
<sup>(3)</sup> $\Omega$ (Achemi- nement)	$A_1$	0	90	1	1	0	1	2	2	2	1
	$A_2$	0	1	4	5	5	40	32	8	4	1
<sup>(4)</sup> $\Omega$ (Rejet Complet)	$\Omega$	25	12	10	8	3	7	5	20	9	1

2. ajouter des zones intermédiaires entre le pays et l'acheminement (niveau département, région, état, etc...);
3. distinguer les villes étrangères des villes du pays traité;
4. distinguer certains groupes de séparateurs dans les cas de distribution spéciale.

En pratique, cette hiérarchie dépend fortement du pays client dans lequel est trié le courrier, des LAP à notre disposition, et du nombre d'exemples de chaque catégorie que nous pouvons obtenir.

Enfin, dans cette 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 en fonction de la décision fournie, nous pouvons particulariser les cas où la vérité était une adresse invalide. Par exemple, lorsqu'un LAP effectue un rejet complet nous avons plus de raisons de penser que la lettre est invalide que lorsqu'il décide une adresse complète. Considérons la matrice de confusion illustrée dans le tableau 5, associé à un classifieur de l'exemple 4.1. Lorsque ce classifieur décide  $\Omega$ , au lieu de convertir cette décision par la fonction de masse vide, nous préférons convertir cette décision par la masse  $m$  définie par :

$$\begin{cases} m(\{inv\}) &= 0.25 \\ m(\Omega) &= 0.75. \end{cases}$$

En effet, lorsque le classifieur a fourni un rejet, dans 25 % des cas l'adresse était totalement invalide, et dans 75 % des cas au moins la ville de la véritable adresse pouvait être précisée. Ainsi même un rejet complet nous apporte une information sur la vérité: nous avons des raisons de croire que la lettre est invalide et des raisons de croire que la ville d'acheminement de cette lettre existe bien.

Sur cette même idée, lorsque le classifieur effectue un rejet distribution et un rejet de pas de porte, il est possible d'affecter des masses à des adresses partiellement invalides. Le détail de cette affectation ne sera pas développé ici.

Remarquons enfin que cette affectation ne produit généralement plus de masses consonantes.

### 5.3. Combinaison des fonctions de croyance

Les LAP étant supposés distincts, et l'affectation fiable, la règle de combinaison conjonctive peut s'appliquer.

#### Exemple 5.8 (Combinaison de 3 LAP)

Considérons 3 LAP, notés  $OCR_1$ ,  $OCR_2$ , et  $OCR_3$ , 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»;

-  $d_3 = \{A_2 R_1 17\}$ , *i.e.*, une décision de niveau distribution et de catégorie «pas de porte».

En notant  $m_i$  la masse issue d'un LAP  $OCR_i$ , 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  $m = m_1 \oplus m_2 \oplus m_3$  donne :

$$\begin{aligned} m(\{A_2R_117\}) &= 0.0845 & m(\{A_1\}) &= 0.0002 \\ m(A_2R_1) &= 0.0067 & m(\Omega) &= 0.0000 \\ m(\{A_2inv\}) &= 0.0002 & m(\emptyset) &= 0.9036. \\ m(A_2) &= 0.0048 \end{aligned} \quad (20)$$

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

## 6. Prise de décision



L'ensemble  $\mathcal{D}$  des décisions pouvant être produites par la combinaison est le même que celui des LAP (cf. exemple 4.1).  $\mathcal{D}$  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  $\Omega$  en un temps raisonnable, ce qui nous empêche de calculer précisément la probabilité pignistique attachée au risque espéré de chaque décision (équation 2). Toutefois, nous pouvons prendre une décision sur un cadre de pari défini comme à un partitionnement ou grossissement de  $\Omega$ , noté  $\Gamma$ , calculé à partir des éléments focaux des masses des LAPs et de la hiérarchie.

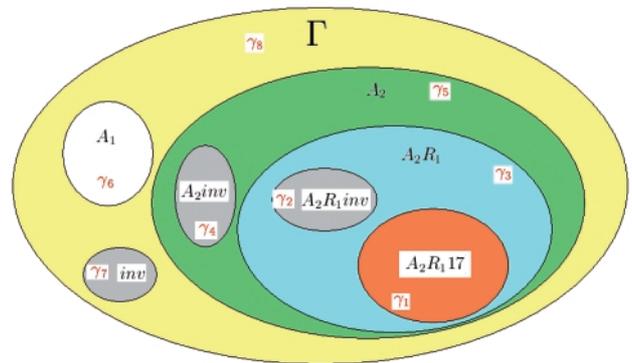


Figure 9. Éléments du cadre de pari  $\Gamma$ , de la hiérarchie (figure 7), et des masses associées à l'exemple 5.8.

### Exemple 6.1 (suite de l'exemple 5.8)

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 «  $A_1$  »,
- $d_3 = A_2$  en faveur du code «  $A_2Rej$  »,
- $d_4 = A_2R_1$  en faveur du code «  $A_2R_1Rej$  »,
- et  $d_5 = A_2R_117$  en faveur du code «  $A_2R_117$  ».

Le grossissement  $\Gamma$  de  $\Omega$  associé à ce cas est illustré sur la figure 9. 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  $\gamma_8 = \Omega \setminus (A_2 \cup \{A_1, inv\})$ .

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.

Tableau 6. Coûts de décisions en fonction de la décision fournie et de la vérité.

Décisions	Vérité	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_5$	$\gamma_6$	$\gamma_7$	$\gamma_8$
$d_1 = \text{« Rej »}$		$C_{Rach}$	$C_{Rach}$	$C_{Rach}$	$C_{Rach}$	$C_{Rach}$	$C_{Rach}$	0	0
$d_2 = \text{« } A_1 \text{ »}$		$C_{Each}$	$C_{Each}$	$C_{Each}$	$C_{Each}$	$C_{Each}$	0	$C_{Each}$	$C_{Each}$
$d_3 = \text{« } A_2Rej \text{ »}$		$C_{Rrue}$	$C_{Rrue}$	$C_{Rrue}$	0	0	$C_{Each}$	$C_{Each}$	$C_{Each}$
$d_4 = \text{« } A_2R_1Rej \text{ »}$		$C_{Rdis}$	0	0	$C_{Erue}$	$C_{Erue}$	$C_{Each}$	$C_{Each}$	$C_{Each}$
$d_5 = \text{« } A_2R_117 \text{ »}$		0	$C_{Edis}$	$C_{Edis}$	$C_{Erue}$	$C_{Erue}$	$C_{Each}$	$C_{Each}$	$C_{Each}$

- Un *coût de rejet* acheminement  $C_{R_{ache}}$  (resp. rue  $C_{R_{rue}}$ , distribution  $C_{R_{dis}}$ ) 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  $C_{E_{ach}}$  (resp. rue  $C_{E_{rue}}$ , distribution  $C_{E_{dis}}$ ) 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 6.2 (suite de l'exemple 6.1)

Les coûts associés à cet exemple sont indiqués dans le tableau 6.

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

$$0 \leq C_{R_{dis}} \leq C_{R_{rue}} \leq C_{R_{ach}} \leq C_{E_{dis}} \leq C_{E_{rue}} \leq C_{E_{ach}}. \quad (21)$$

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 correct, peut avoir un coût moins élevé que le travail demandé par un rejet global.

### Exemple 6.3 (suite de l'exemple 6.2)

À partir des équations (20) et (3), la probabilité pignistique calculée sur  $\Gamma$  est donnée par :

$$\begin{aligned} \text{Bet } P(\{\gamma_1\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + \frac{m(A_2 R_1)}{3} + m(\{A_2 R_1 17\}) \right) \\ &= 0.910 \\ \text{Bet } P(\{\gamma_2\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + \frac{m(A_2 R_1)}{3} + m(\{A_2 R_1 inv\}) \right) \\ &= 0.033 \\ \text{Bet } P(\{\gamma_3\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + \frac{m(A_2 R_1)}{3} \right) \\ &= 0.033 \\ \text{Bet } P(\{\gamma_4\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + \frac{m(A_2)}{5} + m(\{A_2 inv\}) \right) \\ &= 0.012 \\ \text{Bet } P(\{\gamma_5\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + \frac{m(A_2)}{5} \right) \\ &= 0.010 \\ \text{Bet } P(\{\gamma_6\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + m(\{A_1\}) \right) \\ &= 0.002 \\ \text{Bet } P(\{\gamma_7\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} + m(\{inv\}) \right) \\ &= 0.000 \\ \text{Bet } P(\{\gamma_8\}) &= \mathcal{K} \left( \frac{m(\Omega)}{8} \right) \\ &= 0.000 \end{aligned} \quad (22)$$

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

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

$$\begin{aligned} \rho(d_1) &= \sum_{\gamma \in \Gamma} c(d_1, \gamma) \text{Bet } P(\gamma) = \sum_{i \in I \setminus \{7,8\}} C_{R_{ach}} \text{Bet } P(\{\gamma_i\}) \\ &= C_{R_{ach}}, \\ \rho(d_2) &= \sum_{\gamma \in \Gamma} c(d_2, \gamma) \text{Bet } P(\gamma) = \sum_{i \in I \setminus \{6\}} C_{E_{ach}} \text{Bet } P(\{\gamma_i\}) \\ &= 0.998 C_{E_{ach}}, \\ \rho(d_3) &= \sum_{i \in \{1,2,3\}} C_{R_{rue}} \text{Bet } P(\{\gamma_i\}) + \sum_{i \in \{6,7,8\}} C_{E_{ach}} \text{Bet } P(\{\gamma_i\}) \\ &= 0.976 C_{R_{rue}} + 0.002 C_{E_{ach}}, \\ \rho(d_4) &= C_{R_{dis}} \text{Bet } P(\{\gamma_1\}) + \sum_{i \in \{4,5\}} C_{E_{rue}} \text{Bet } P(\{\gamma_i\}) \\ &\quad + \sum_{i \in \{6,7,8\}} C_{E_{ach}} \text{Bet } P(\{\gamma_i\}) \\ &= 0.910 C_{R_{dis}} + 0.022 C_{E_{rue}} + 0.002 C_{E_{ach}}, \\ \rho(d_5) &= \sum_{i \in \{2,3\}} C_{E_{dis}} \text{Bet } P(\{\gamma_i\}) + \sum_{i \in \{4,5\}} C_{E_{rue}} \text{Bet } P(\{\gamma_i\}) \\ &\quad + \sum_{i \in \{6,7,8\}} C_{E_{ach}} \text{Bet } P(\{\gamma_i\}) \\ &= 0.066 C_{E_{dis}} + 0.022 C_{E_{rue}} + 0.002 C_{E_{ach}}. \end{aligned}$$

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 (21). 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; autrement dit, plus les coûts d'erreur seront élevés par rapport aux coûts de rejet, plus il faudra une probabilité élevée sur un élément différent d'un rejet pour pouvoir le décider. Par exemple, avec le 6-uplet  $(C_{R_{dis}}, C_{R_{rue}}, C_{R_{ach}}, C_{E_{dis}}, C_{E_{rue}}, C_{E_{ach}})$  égal à :

- (1,2,3,4,5,6), la décision choisie est  $d_5$  i.e. «  $A_2 R_1 17$  » ;
- (1,2,3,20,25,50), la décision choisie est  $d_4$  i.e. «  $A_2 R_1 Rej$  » ;
- (1,2,3,40,50,60), la décision choisie est  $d_3$  i.e. «  $A_2 Rej$  » ;
- (1,2,3,400,500,600), la décision choisie est  $d_1$  i.e. «  $Rej$  ».

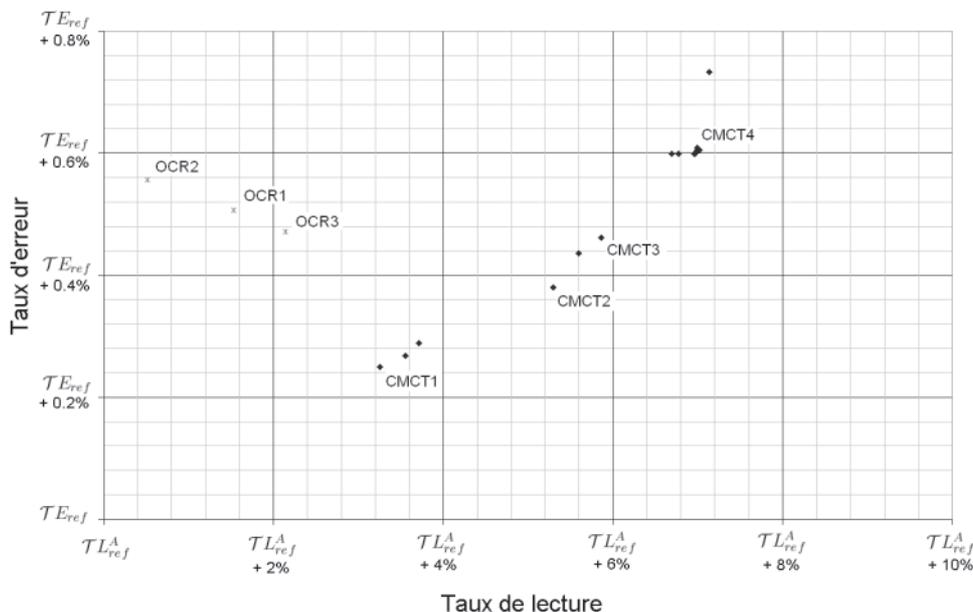
Idéalement ces coûts sont donnés par des experts du domaine ou le client, et reflètent des coûts financiers réels. Mais si le client préfère une combinaison avec un taux d'erreur contrôlé à certains niveaux, ces coûts peuvent être déterminés à partir d'un ensemble d'apprentissage de manière à obtenir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

## 7. Résultats expérimentaux

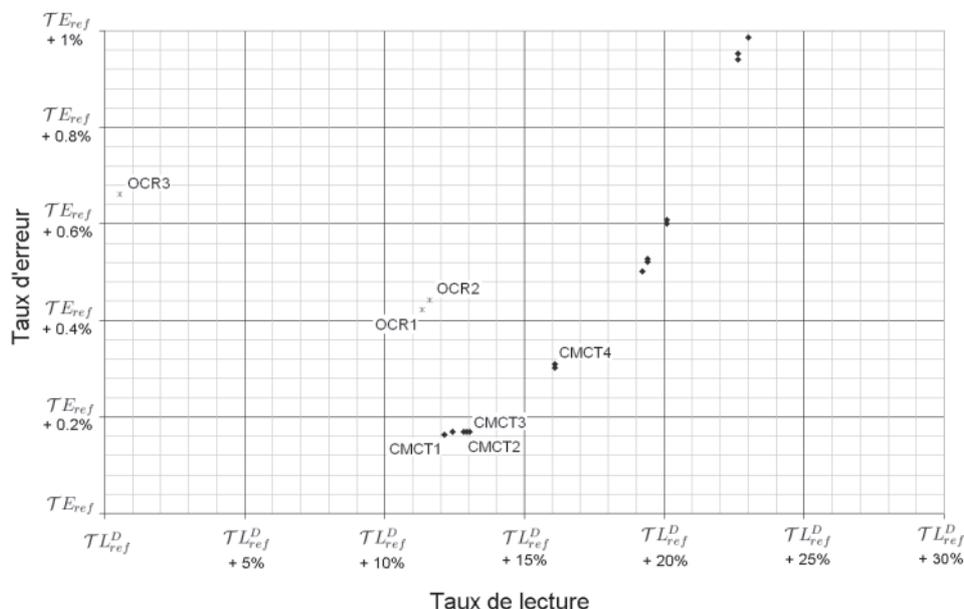
Dans cette application, nous illustrons notre démarche de combinaison en considérant trois LAP  $OCR_1$ ,  $OCR_2$  et  $OCR_3$  fournissant des décisions en accord avec la hiérarchie (figure 7).

### 7.1. Choix des coûts

Les coûts sont déterminés manuellement de façon à obtenir le meilleur taux de lecture possible en maintenant le taux d'erreur



(a) Performances au niveau acheminement.



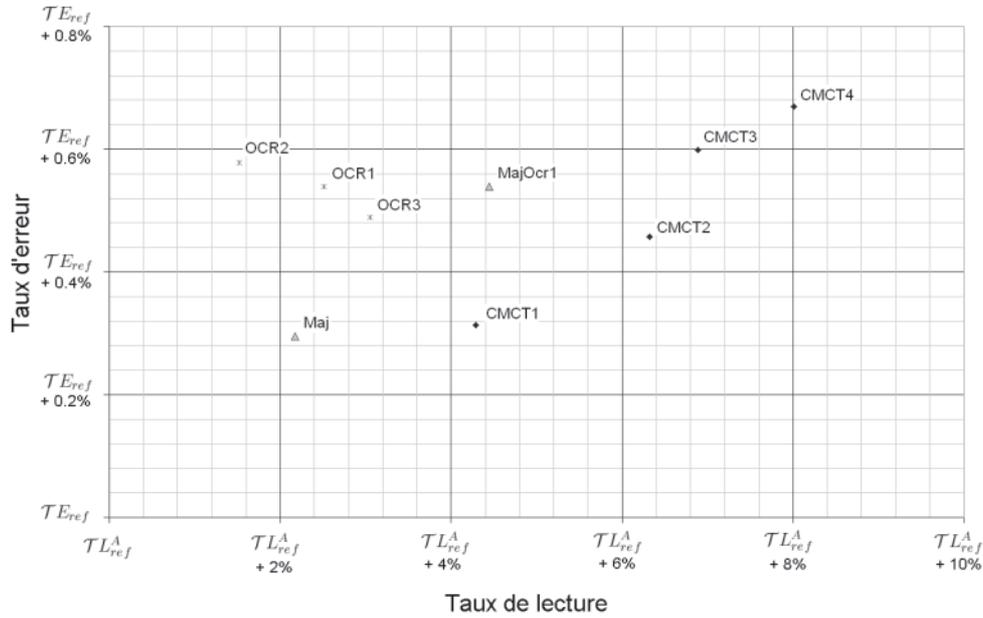
(b) Performances au niveau distribution.

Figure 10. Performances des LAP et de différents points de fonctionnement de la combinaison CMCT réalisées sur l'ensemble d'apprentissage.

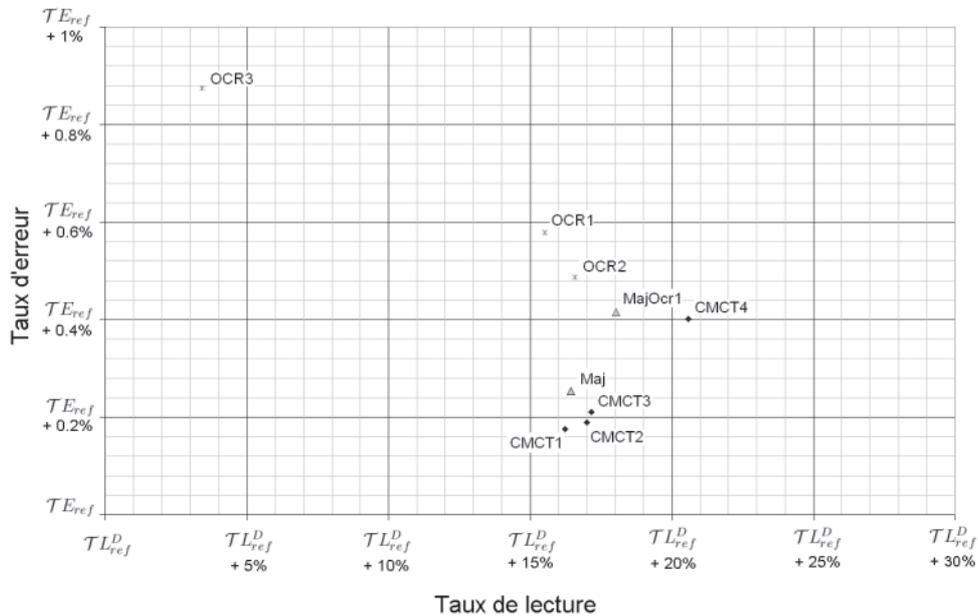
dans un intervalle jugé acceptable. Pour cela, nous disposons d'un ensemble d'apprentissage composé d'environ 25 000 lettres. C'est sur ce lot d'images que les matrices de confusion ont été calculées, et les coûts ajustés. Un autre lot de 25 000 lettres constitue l'ensemble de test sur lequel ont été 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  $TL_{ref}^A$  (resp.  $TL_{ref}^D$ ).



(a) Performances au niveau acheminement.



(b) Performances au niveau distribution.

Figure 11. 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.

De même, le taux d'erreur sera exprimé en fonction d'un taux de référence en erreur, noté  $T_{E_{ref}}$ .

La figure 10(a) illustre des performances obtenues en phase d'apprentissage au niveau acheminement, tandis que la figure 10(b) illustre des performances obtenues au niveau distribution. Sur ces deux figures, sont représentés différents points de fonctionnement, c'est-à-dire des couples (taux de lecture, taux

d'erreur), de la combinaison implémentée à partir du modèle défini dans cet article, nommée CMCT pour *Combinaison basée sur le MCT*. Ces points de fonctionnement sont obtenus à partir de différents coûts de décision. Nous pouvons remarquer certains « sauts » dans les performances de la combinaison. Ceux-ci proviennent de l'affectation qui caractérise, pour chaque LAP, tout lot de courrier de même type de décision de manière

identique. Il faudrait tenir compte de nouvelles informations pour discriminer des lettres à l'intérieur de ces lots. Nous donnons des éléments dans les perspectives.

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.

## 7.2. Schémas à base de votes

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 [11]. 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  $Ocr_1$ . Cette méthode sera notée *MajOCR<sub>1</sub>*.

### Exemple 7.1

Si le LAP  $OCR_1$  décide «*A<sub>2</sub>Rej*», le LAP  $OCR_2$  décide «*Rej*» et le LAP  $OCR_3$  décide «*Rej*», alors le processus *Maj* décidera «*Rej*» et le processus *MajOCR<sub>1</sub>* décidera «*A<sub>2</sub>Rej*».

## 7.3. 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 11(a) et 11(b).

Aux niveaux acheminement et distribution, un gain notable est obtenu par rapport aux performances individuelles des LAP et des schémas à base de votes.

Dans les deux cas, la combinaison CMCT démontre une robustesse dans son comportement : les points de fonctionnement restent dans la même «zone de performance» sur les deux lots de courrier.

# 8. Conclusion et perspectives

Dans cet article, un modèle de fusion d'adresses postales basées sur la théorie des fonctions de croyance a été présenté. 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. 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 l'affectation Bayésienne et celle de Xu *et al.* Du fait de la construction des masses uniquement à partir des décisions, ce modèle de combinaison peut s'appliquer à tout type de classifieurs. Bien que n'employant que peu d'outils dans la manipulation des fonctions de croyance, ce modèle fournit déjà des performances intéressantes. La prise en compte de nouvelles informations offre de nombreuses perspectives. Par exemple, le modèle ne tient pas compte des scores de confiance fournis par certains LAP en plus de leur décision. De même, un travail sur la prise en compte des dépendances entre LAPs dans la fusion est en cours [7]. Ces raffinements du modèle devraient encore améliorer la fusion proposée.

## Références

- [1] A. APPRIOU, Perspectives liées à la fusion de données, *Science et Défense* 90, 1990.
- [2] B. BENYOUB, Une application industrielle de reconnaissance d'adresses, *Colloque National sur l'Écrit et le Document*, pages 93-100, Nantes, France, Juillet 1996.
- [3] I. BLOCH, *Fusion d'informations en traitement du signal et des images*, Hermès, Paris, France, 2003.
- [4] G. DE LEO, M. VICENZI et C. FRANZONE. ELSAG SPA. European patent: Mail recognition method. Number: EP 1 594 077 A2, November 2005.
- [5] M.H. DEGROOT, *Optimal Statistical decisions*, McGraw-Hill, New York, 1970.
- [6] T. DENÈUX, Analysis of evidence-theoretic decision rules for pattern classification, *Pattern Recognition*, 30(7): 1095-1107, 1997.
- [7] T. DENÈUX, The cautious rule of combination for belief functions and some extensions, *Proceedings of the 9<sup>th</sup> International Conference on Information Fusion*, Florence, Italy, July 2006.
- [8] 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.
- [9] M. FISHER, Siemens AG. International patent: System and method for smart polling. Number: WO 2005/050545 A1, June 2005.
- [10] 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.
- [11] 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.
- [12] D. MERCIER, T. DENÈUX et M. MASSON, General correction mechanisms for weakening or reinforcing belief functions. *Proceedings of the 9<sup>th</sup> International Conference on Information Fusion, FUSION'2006*, Florence, Italie, paper 146, July 10-14 2006.
- [13] D. MERCIER, B. QUOST, et T. DENÈUX, Contextual discounting of belief functions. Dans Lluís Godo (Eds.), *Proceedings of the 8<sup>th</sup> European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 552-562, Barcelona, Spain, July 6-8 2005. Springer Verlag.
- [14] L.J. SAVAGE, *The Foundations of Statistics*, Wiley, New York, 1954.
- [15] G. SHAFER, *A mathematical theory of evidence*, Princeton University Press, Princeton, N.J., 1976.
- [16] 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.

- [17] Ph. SMETS, Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem, *International Journal of Approximate Reasoning*, 9: 1-35, 1993.
- [18] 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.
- [19] 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.
- [20] Ph. SMETS, Data fusion in the transferable belief model, *Proceedings of the 3rd International Conference on Information Fusion, FUSION 2000*, pages PS21-33, Paris, France, 2000.

- [21] Ph. SMETS, Decision making in the TBM: the necessity of the pignistic transformation, *International Journal of Approximate Reasoning*, 38(2): 133-147, 2005.
- [22] Ph. SMETS, Analyzing the combination of conflicting belief functions, *Disponible en ligne <http://iridia.ulb.ac.be/psmets/CombiConfl.pdf>*. À paraître dans *Information Fusion*, 2006.
- [23] Ph. SMETS et R. KENNES, The Transferable Belief Model, *Artificial Intelligence*, 66: 191-243, 1994.
- [24] Traitement du signal (Eds.), Numéro special: Fusion de données, *Traitement du signal*, 14(5), 1997.
- [25] 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.



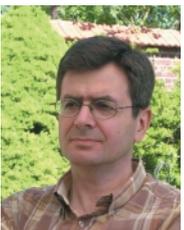
David Mercier

D. Mercier est titulaire d'un DEA RCFR (Intelligence Artificielle) obtenu à l'Université Paul Sabatier de Toulouse en 2003. Il est doctorant à l'Université de Technologie de Compiègne, et effectue une thèse CIFRE dans l'entreprise Solystic, où il étudie la problématique de fusion d'adresses postales dans le cadre du modèle des croyances transférables.



Geneviève Cron

Après sa thèse sur le diagnostic en reconnaissance des formes, fusion et logique floue, soutenue en 1999 à l'Université de Technologie de Compiègne, G. Cron entre chez Solystic, fabricant de machines de tri postal. Elle y étudie les problématiques de combinaison d'information dans l'équipe de logiciel de reconnaissance d'adresse postale sur le courrier dactylographié et manuscrit.



Thierry Dencœur

T. Dencœur est ingénieur civil (1985) et docteur (1989) de l'Ecole Nationale des Ponts et Chaussées. Il est actuellement professeur à l'Université de Technologie de Compiègne, co-responsable du thème Apprentissage Statistique, Reconnaissance de formes, Image et Décision au sein du laboratoire Heudiasyc (UMR CNRS 6599), et rédacteur en chef de la revue *International Journal of Approximate Reasoning* publiée par Elsevier. Ses recherches portent principalement sur la théorie des fonctions de croyance et son application à la représentation des incertitudes en reconnaissance de formes et en fusion d'informations.



Marie-Hélène Masson

M.-H. Masson est ingénieur et docteur HDR de l'Université de Technologie de Compiègne. Elle est Maître de Conférences à l'Université de Picardie Jules Verne et membre du laboratoire Heudiasyc de l'Université de Technologie de Compiègne. Ses recherches portent sur l'analyse et la fusion de données par des théories non probabilistes de gestion de l'incertain (théorie des possibilités, théorie des fonctions de croyance).

