

Diagnostic par Reconnaissance de formes :  
Approches non probabilistes

Thierry Dencœux

31 janvier 2001

## Table des matières

3	Diagnostic par Reconnaissance de formes : Approches non probabilistes	3
3.1.	Introduction . . . . .	3
3.2.	Approche floue . . . . .	7
3.2.1.	Notions préliminaires . . . . .	7
3.2.2.	Classification floue non supervisée . . . . .	8
3.2.3.	Définition des fonctions d'appartenance . . . . .	13
3.2.4.	Décision à partir des degrés d'appartenance . . . . .	15
3.2.5.	Compléments bibliographiques . . . . .	18
3.3.	Approche crédibiliste . . . . .	19
3.3.1.	Cadre général . . . . .	19
3.3.2.	Théorie des fonctions de croyance . . . . .	21
3.3.3.	Construction des fonctions de croyance en discrimination	22
3.3.4.	Principes de décision . . . . .	26
3.3.5.	Extensions . . . . .	29
3.4.	Conclusion . . . . .	32

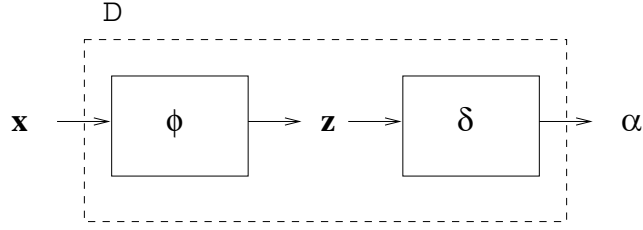
## Chapitre 3

# Diagnostic par Reconnaissance de formes : Approches non probabilistes

### 3.1. Introduction

On considère dans ce chapitre un système quelconque  $S$  caractérisé à chaque instant par une variable  $y$  décrivant son *mode de fonctionnement*, et par un vecteur de mesures  $\mathbf{x} \in \mathfrak{R}^d$ . Dans l'approche du diagnostic dite « par Reconnaissance de Formes » (cf. Chapitre 2), la connaissance disponible est supposée limitée à un ensemble d'observations (éventuellement imprécises ou incertaines) des variables  $\mathbf{x}$  et  $y$ , réalisées sur le système à différents instants. Le problème consiste alors à exploiter cette information statistique dans le but de prédire (lorsque cela est jugé possible) la valeur de la variable  $y$  pour toute nouvelle valeur de la variable  $\mathbf{x}$ , appelée « vecteur forme ». Ce problème peut être formalisé comme un problème de décision dans l'incertain, c'est-à-dire comme le choix d'une action parmi un ensemble  $\mathcal{A}$ , sur la base de connaissances partielles.

En diagnostic, l'ensemble  $\mathcal{A}$  inclut typiquement des actions d'affectation (consistant à affirmer que le système se trouve dans un certain mode de fonctionnement), et des actions de rejet, ou d'« abstention » qui peuvent être choisies lorsque l'incertitude est jugée trop grande. Cette incertitude pouvant revêtir différentes formes, il peut être judicieux de distinguer entre plusieurs actions de rejet auxquelles seront associées des interprétations différentes. Ainsi, Dubuisson [DUB 90], [DUB 93] a introduit une distinction importante entre les notions de *rejet d'ambiguïté* et *rejet de distance*. Dans le premier cas, on se refuse à faire une prédiction car plusieurs modes semblent également plausibles ; l'ambiguïté peut alors parfois être levée en réalisant de nouvelles mesures sur le système. Dans le second cas, l'incertitude provient du caractère atypique ou « aberrant » des observations, qui rend peu vraisemblable l'appartenance du système



**Figure 3.1.** Composantes d'un système de diagnostic par RdF.

à aucun des modes connus. Une méthodologie de décision susceptible d'établir une telle distinction doit nécessairement prendre en compte, entre l'espace des observations et celui des décisions, un niveau intermédiaire de *description des incertitudes* sur le mode de fonctionnement du système [DEN 98b].

Soit  $D : \mathbb{R}^d \rightarrow \mathcal{A}$  la fonction, appelée *fonction de décision*, qui associe une action à chaque vecteur forme  $\mathbf{x}$ . Dans les approches probabilistes décrites au chapitre précédent,  $y$  est supposée être une variable qualitative à  $M$  modalités :  $y \in \Omega = \{\omega_k\}_{k=1}^M$ , et l'incertitude sur l'état du système est modélisée par une distribution de probabilité *a posteriori*  $\Pr(\cdot/\mathbf{x})$  sur  $\Omega$ , pouvant être représentée par un vecteur  $\mathbf{z} = \{\Pr(\omega_k/\mathbf{x})\}_{k=1}^M \in [0, 1]^M$  vérifiant

$$\sum_{k=1}^M z_k = 1. \quad [3.1]$$

Une stratégie de décision (basée, par exemple, sur le principe de minimisation du coût espéré) associe ensuite une action  $\alpha \in \mathcal{A}$  à chaque valeur de  $\mathbf{z}$ . La fonction de décision peut donc se décomposer en

$$D = \delta \circ \phi, \quad [3.2]$$

où  $\phi : \mathbf{x} \mapsto \mathbf{z}$  traduit un mécanisme de quantification de l'incertitude sur l'état du système, et où  $\delta : \mathbf{z} \mapsto \alpha$  modélise une stratégie de décision en environnement incertain (Figure 3.1).

L'approche probabiliste des problèmes de décision (cf. Chapitre 2) a le mérite d'être à la fois bien justifiée théoriquement, et relativement simple d'un point de vue calculatoire. Néanmoins, elle possède comme toute méthode un domaine de validité délimité par certaines hypothèses, qui sont ici principalement au nombre de deux.

La première de ces hypothèses consiste à considérer la variable  $y$  comme monovaluée, ce qui revient à considérer que le mode de fonctionnement du système à chaque instant peut être défini de manière unique et non ambiguë.

Cette hypothèse est en général bien adaptée à des problèmes tels que le diagnostic de pannes dans des systèmes technologiques relativement simples : par exemple, la rupture d'un élément mécanique ou la défaillance d'un composant électronique est un fait qui, lorsqu'il se réalise, place objectivement et immédiatement le système considéré dans un certain état. En revanche, la notion d'état ou de mode de fonctionnement devient moins nette lorsque l'on admet dans la modélisation la possibilité de passage *graduel* d'un état à un autre, ou lorsque la notion d'état sert à qualifier de manière imprécise le comportement d'un système complexe. Par exemple, le problème du suivi de l'usure d'un outil de coupe peut être formalisé en considérant que le système évolue progressivement d'un état initial vers un état final où le fonctionnement de l'outil n'est plus acceptable [ZIE 94], [ZIE 95]. Dans un tel modèle, la proposition « l'outil est usé » n'est plus un événement au sens probabiliste du terme puisqu'on ne peut lui associer une valeur de vérité binaire. Un autre exemple d'application dans laquelle la notion de mode de fonctionnement est intrinsèquement imprécise concerne le diagnostic d'hypovigilance d'un conducteur automobile [PEL 93c], [PEL 93a]. Dans ce cas, c'est principalement la complexité du système étudié (un être humain) qui motive sa description par des concepts imprécis, mais néanmoins pertinents dans le cadre de l'application envisagée, tels que ceux de veille et d'hypovigilance.

La théorie des ensembles flous [ZAD 65], [KLI 95], [BOU 95] s'est imposée ces dernières années comme un formalisme adapté à la modélisation et à la manipulation d'informations vagues et imprécises. Schématiquement, un *sous-ensemble flou*  $F$  d'un ensemble  $A$  peut être défini comme une partie de  $A$  aux frontières mal définies, chaque élément  $u$  de  $A$  ayant un degré d'appartenance  $\mu_F(u)$  à  $F$ . Une application de cette notion en diagnostic consiste à supposer que la variable  $y$  décrivant le mode de fonctionnement du système ne prend pas une seule valeur, mais simultanément plusieurs valeurs à des degrés différents. Formellement, le domaine de  $y$  est donc l'ensemble  $[0, 1]^\Omega$  des sous-ensembles flous de  $\Omega$ . Par exemple, la proposition  $y = F$  où  $F \in [0, 1]^\Omega$  signifie que le système se trouve dans chaque mode  $\omega \in \Omega$  avec un degré  $\mu_F(\omega) \in [0, 1]$ . La quantité  $\mu_F(\omega)$  est appelée degré d'appartenance au mode  $\omega$ . Il est quelquefois judicieux d'imposer la condition

$$\sum_{k=1}^M \mu_F(\omega_k) = 1 \quad [3.3]$$

qui généralise la notion d'unicité du mode de fonctionnement ( $F$  est alors un « singleton flou » de  $\Omega$ ). Pour chaque valeur  $\mathbf{x}$  du vecteur d'observations, il s'agit alors de définir un vecteur d'appartenance  $\mathbf{z} = \{\mu_k(\mathbf{x})\}_{k=1}^M \in [0, 1]^M$ ,  $\mu_k(\mathbf{x})$  désignant une évaluation du degré d'appartenance du système au mode de fonctionnement  $\omega_k$ . La fonction  $\mu_k : \mathbf{x} \mapsto \mu_k(\mathbf{x})$  est appelée *fonction d'appartenance* associée au mode  $\omega_k$ .

Remarquons ici que, bien que ce formalisme puisse paraître assez proche du modèle probabiliste d'un point de vue purement formel, il s'en distingue fondamentalement par l'interprétation des concepts manipulés et, en conséquence, par les traitements ultérieurs effectués sur les nombres  $z_k$ . La construction de fonctions d'appartenance à partir d'un ensemble de données étiquetées ou non, ainsi que leur utilisation en diagnostic, feront l'objet du paragraphe 3.2 de ce chapitre.

La seconde hypothèse à la base des méthodes probabilistes de reconnaissance des formes porte sur l'existence d'une loi de probabilité  $\Pr(\cdot/\mathbf{x})$  sur  $\Omega$ . Dans un cadre fréquentiste, on postule habituellement l'existence de lois conditionnelles des observations  $p(\mathbf{x}|y)$  relatives à chaque mode de fonctionnement, et d'une loi *a priori*  $\Pr(y)$  sur  $\Omega$ . Les probabilités *a posteriori*  $\Pr(\cdot/\mathbf{x})$  sont alors déduites par la règle de Bayes. Ces probabilités théoriques, inconnues, sont couramment estimées à partir d'un ensemble d'observations  $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$  supposé issu d'un échantillon indépendant et identiquement distribué des variables aléatoires  $\mathbf{x}$  et  $y$  [DUD 73], [FUK 90]. Remarquons que, si ce cadre théorique est couramment accepté, il n'en constitue pas moins un modèle simplifié de la réalité, dont la pertinence est parfois discutable. Par exemple, il n'est pas rare qu'un système de décision soit élaboré à partir de données expérimentales ou simulées n'ayant pas les mêmes propriétés statistiques que les données issues de l'environnement opérationnel du système, pouvant lui même varier au cours du temps.

Dans une perspective différente, l'existence d'une loi de probabilité subjective sur  $y$  conditionnellement à  $\mathbf{x}$  peut être justifiée si l'on adopte le cadre Bayésien pour la représentation des connaissances partielles. L'existence d'une loi de probabilité sur les états du système découle alors non plus de propriétés physiques de ce système, mais de principes de rationalité supposés guider le comportement d'un décideur en situation d'incertitude. Sans entrer ici dans une discussion approfondie des présupposés de cette approche, nous nous contenterons de mentionner qu'elle est depuis quelques années mise en question [DUB 88], [SME 98a], et concurrencée par des formalismes non probabilistes tels que la théorie des possibilités, ou la théorie des croyances dont l'application en reconnaissance des formes s'est révélée très féconde. Dans cette dernière approche, l'incertitude sur l'état du système à un instant donné est supposée pouvoir être décrite par une fonction  $m$  affectant une masse  $m(A)$  à chaque partie  $A$  de  $\Omega$ . La fonction  $m$  peut donc être représentée par un vecteur  $\mathbf{z} = \{m(A)\}_{A \subseteq \Omega}$  à  $2^M$  composantes. La construction par apprentissage de la fonction de masse  $m$ , ainsi que son utilisation en décision, feront l'objet du paragraphe 3.3.

## 3.2. Approche floue

### 3.2.1. Notions préliminaires

La théorie des ensembles flous (TEF) a été introduite dans le but de faciliter l'analyse de systèmes complexes en autorisant la représentation et le traitement de connaissances imprécises [YAG 87]. Le concept central de la TEF est celui d'ensemble flou. Un sous-ensemble classique  $A$  d'un univers  $\Omega$  peut être décrit par sa fonction d'appartenance  $\mu_A : \Omega \mapsto \{0, 1\}$  vérifiant

$$\mu_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} \quad [3.4]$$

De même, il est possible de définir formellement un sous-ensemble flou  $F$  de  $\Omega$  comme un « sous-ensemble généralisé » de  $\Omega$  caractérisé par une fonction d'appartenance  $\mu_F : \Omega \mapsto [0, 1]$ . Chaque quantité  $\mu_F(\omega)$  est interprétée comme le degré d'appartenance de l'élément  $\omega$  à  $F$ . Le support de  $F$  est défini comme l'ensemble classique des éléments de  $\Omega$  ayant un degré d'appartenance non nul à  $F$ , tandis que le noyau de  $F$  est l'ensemble classique des éléments  $\omega$  appartenant totalement à  $F$ , c'est-à-dire tels que  $\mu_F(\omega) = 1$ .

Au-delà de la caractérisation formelle des ensembles flous, quelle signification convient-il d'attacher à ce concept ? Sans entrer ici dans les détails d'un débat assez complexe, il nous semble utile de mentionner deux interprétations couramment proposées [RUS 98b].

La première est liée à la notion de distance (ou similarité) à un ou plusieurs éléments de référence ou *prototypes*. Par exemple, l'ensemble des nombres « environ égaux à trois » peut être défini comme un ensemble flou dont la fonction d'appartenance est une mesure de similarité au nombre 3.

La seconde interprétation est associée à la notion de « contrainte flexible » associée à un certain problème. Par exemple, une proposition telle que « La pression est légèrement supérieure à trois bars » restreint l'ensemble des valeurs possibles de la pression sans éliminer catégoriquement des valeurs telles que 3,5 ou 4 bars ; il est donc possible de lui associer un ensemble flou, dont la fonction d'appartenance traduira le degré de satisfaction de la contrainte.

L'un des intérêts essentiels de la notion d'ensemble flou réside dans la possibilité d'étendre les opérateurs ensemblistes usuels tels que l'intersection, l'union et la complémentation. Si  $F$  et  $G$  sont deux sous-ensembles flous de  $\Omega$ , on pose généralement :

$$\mu_{F \cap G}(\omega) = \min(\mu_F(\omega), \mu_G(\omega)) \quad [3.5]$$

$$\mu_{F \cup G}(\omega) = \max(\mu_F(\omega), \mu_G(\omega)) \quad [3.6]$$

$$\mu_{\overline{F}}(\omega) = 1 - \mu_F(\omega) \quad [3.7]$$

pour tout  $\omega \in \Omega$ ,  $\overline{F}$  désignant le complémentaire de  $F$  dans  $\Omega$ .

Cette extension, initialement proposée par Zadeh [ZAD 65], n'est cependant pas unique. Plus généralement, posons :

$$\mu_{F \cap G}(\omega) = \top(\mu_F(\omega), \mu_G(\omega)) \quad [3.8]$$

où  $\top$  est une application de  $[0, 1] \times [0, 1]$  dans  $[0, 1]$ . Il est naturel d'imposer à  $\top$  les propriétés suivantes, pour tout  $u, v, w \in [0, 1]$  :

1.  $v \leq w \Rightarrow \top(u, v) \leq \top(u, w)$  [isotonie] ;
2.  $\top(u, v) = \top(v, u)$  [commutativité] ;
3.  $\top(u, \top(v, w)) = \top(\top(u, v), w)$  [associativité] ;
4.  $\top(u, 1) = u$  [élément neutre 1].

Ces conditions définissent une norme triangulaire ou *t-norme*. Il est facile de vérifier que l'opérateur min est une *t-norme*, de même que, par exemple, le produit algébrique  $\top(u, v) = uv$ .

De façon similaire, l'union peut être définie par une relation du type :

$$\mu_{F \cup G}(\omega) = \perp(\mu_F(\omega), \mu_G(\omega)) \quad [3.9]$$

où  $\perp$  est une conorme triangulaire ou *t-conorme*, c'est-à-dire une application de  $[0, 1] \times [0, 1]$  dans  $[0, 1]$  vérifiant les propriétés de monotonie, commutativité et associativité, et admettant 0 pour élément neutre :

$$\perp(u, 0) = u \quad \forall u \in [0, 1] \quad [3.10]$$

Le maximum et la somme algébrique  $\perp(u, v) = u + v - uv$  sont deux exemples de *t-conormes*.

Une *t-norme* et une *t-conorme* triangulaires sont dites duales pour l'opération de complément standard si elles satisfont aux lois de De Morgan :

$$1 - \perp(u, v) = \top(1 - u, 1 - v) \quad [3.11]$$

et

$$1 - \top(u, v) = \perp(1 - u, 1 - v) \quad [3.12]$$

pour tout  $u, v \in [0, 1]$ . On montre facilement que les couples  $(\min(u, v), \max(u, v))$  et  $(uv, u + v - uv)$  vérifient cette propriété de dualité.

### 3.2.2. Classification floue non supervisée

#### Notion de partition floue

Le problème central en reconnaissance de formes floue consiste en l'apprentissage des fonctions d'appartenance à partir des données. Lorsque celles-ci se



composent de  $N$  vecteurs  $\{\mathbf{x}_i\}_{i=1}^N$  non étiquetés (c'est-à-dire de  $N$  observations de la variable  $\mathbf{x}$  auxquelles n'est associée aucune observations de la variable  $y$ ), l'apprentissage est dit *non supervisé*. Ce problème rejoint celui de la classification automatique en Analyse des Données, et peut être formalisé comme la recherche d'une partition floue des observations, généralisant la notion classique de partition d'un ensemble de points [DUN 74], [BEZ 74].

Une partition de  $N$  vecteurs en  $M$  classes peut être décrite par une matrice  $U$  à  $N$  lignes et  $M$  colonnes, de terme général  $u_{i,k}$ , et telle que  $u_{i,k} = 1$  si  $\mathbf{x}_i$  appartient à la classe  $k$ , et  $u_{i,k} = 0$  sinon. D'après la définition d'une partition, tout vecteur  $\mathbf{x}_i$  appartient à une et une seule classe, ce qui implique que

$$\sum_{k=1}^M u_{i,k} = 1 \quad \forall i \quad [3.13]$$

Par ailleurs, on peut supposer qu'aucune classe n'est vide, ce qui implique

$$\sum_{i=1}^N u_{i,k} > 0 \quad \forall k \quad [3.14]$$

Une matrice de partition floue se définit comme une matrice  $U$  de terme général  $u_{i,k} \in [0, 1]$  vérifiant les équations [3.13] et [3.14] précédentes. La quantité  $u_{i,k}$  s'interprète alors comme un degré d'appartenance (prenant n'importe quelle valeur réelle entre 0 et 1) de l'objet  $i$  à la classe  $k$ .

### Algorithmes HCM et FCM

L'un des premiers algorithmes proposés pour construire automatiquement une partition floue d'un ensemble de vecteurs de  $\mathfrak{R}^d$  est l'algorithme FCM (*Fuzzy c-means*), ou algorithme des centres mobiles flou, introduit par Dunn [DUN 74] et Bezdek [BEZ 74], et étudié de manière très approfondie par Bezdek [BEZ 81], [BEZ 92]. Cette méthode généralise l'algorithme des centres mobiles [THO 53] (ou algorithme HCM : *hard c-means*), qui est à la base de la méthode des nuées dynamiques [DID 74], [CEL 89].

Les algorithmes HCM et FCM consistent à minimiser itérativement un critère de la forme :

$$J_m(U, V) = \sum_{k=1}^M \sum_{i=1}^N u_{i,k}^m \|\mathbf{x}_i - \mathbf{v}_k\|_A^2 \quad [3.15]$$

où  $\mathbf{v}_k$  est le prototype ou centre de la classe  $k$ ,  $m \geq 1$  est une constante,  $V = (\mathbf{v}_1, \dots, \mathbf{v}_M)$ , et  $\|\cdot\|_A$  désigne la distance euclidienne généralisée associée

à une matrice symétrique définie positive  $A$  :

$$\|x_i - \mathbf{v}_k\|_A^2 = (\mathbf{x}_i - \mathbf{v}_k)^T A (\mathbf{x}_i - \mathbf{v}_k) \quad [3.16]$$

Remarquons que  $J_1$  n'est autre que le critère d'inertie intraclasse du nuage des points affectés chacun d'une masse unité. La matrice  $A$  est souvent prise égale à l'identité (on retrouve alors la distance euclidienne classique), ou à l'inverse de la matrice  $\widehat{\Sigma}$  de covariance empirique des données,  $\|\mathbf{x}_i - \mathbf{v}_k\|_A^2$  n'étant autre dans ce cas que la distance de Mahalanobis estimée du vecteur  $\mathbf{x}_i$  à la classe  $k$ .

Dans le cas où  $m = 1$  et où l'on se restreint à la recherche d'une partition classique ( $u_{i,k} \in \{0, 1\}$ ), une condition nécessaire pour que  $(U, V)$  soit un minimum local de  $J_1$  est que l'on ait :

$$u_{i,k} = \begin{cases} 1 & \|\mathbf{x}_i - \mathbf{v}_k\|_A \leq \|\mathbf{x}_i - \mathbf{v}_\ell\|_A \quad \forall \ell \neq k \\ 0 & \text{sinon} \end{cases} \quad \forall i, k \quad [3.17]$$

$\mathbf{v}_k$  est alors le centre de gravité du nuage de points correspondant à la classe  $k$  :

$$\mathbf{v}_k = \frac{\sum_{i=1}^N u_{i,k} \mathbf{x}_i}{\sum_{i=1}^N u_{i,k}} \quad \forall k \quad [3.18]$$

Dans le cas où  $m > 1$  et où  $U$  est une matrice de partition floue, les conditions nécessaires pour que  $(U, V)$  soit un minimum de  $J_m$  deviennent :

$$u_{i,k} = \left[ \sum_{\ell=1}^M (\|x_i - \mathbf{v}_k\|_A / \|x_i - \mathbf{v}_\ell\|_A)^{2/(m-1)} \right]^{-1} \quad \forall i, k \quad [3.19]$$

$$\mathbf{v}_k = \frac{\sum_{i=1}^N u_{i,k}^m \mathbf{x}_i}{\sum_{i=1}^N u_{i,k}^m} \quad \forall k \quad [3.20]$$

Les algorithmes HCM et FCM sont des méthodes d'optimisation alternée des critères  $J_1$  pour le premier, et  $J_m$  ( $m > 1$ ) pour le second. Leur principe consiste à mettre à jour alternativement la partition par les équations [3.17] ou [3.19], et les prototypes par les équations [3.18] ou [3.20]. L'algorithme général est décrit de manière plus détaillée dans le tableau 3.1.

Bezdek [BEZ 80], [BEZ 87] a montré la convergence de l'algorithme FCM vers un minimum local ou un point selle du critère  $J_m$ .

**Tableau 3.1.** *Algorithmes HCM et FCM.*

Données	$N$ vecteurs non étiquetés $\{\mathbf{x}_i\}_{i=1}^N$
Paramétrage	nombre de classes $M$ nombre maximal d'itérations $T$ exposant $m \geq 1$ norme pour le calcul des distances $\ \cdot\ _A$ seuil d'arrêt $\epsilon$
Initialisation	tirage au hasard des prototypes $V_0 = (\mathbf{v}_1^{(0)}, \dots, \mathbf{v}_M^{(0)})$ . $t \leftarrow 0$
Itérations	<b>Répéter</b> $t \leftarrow t + 1$ calcul de la partition $U_t$ par [3.17] ou [3.19] calcul des prototypes $V_t$ par [3.18] ou [3.20] <b>Tant que</b> ( $t \leq T$ ) et $\ V_t - V_{t-1}\  \geq \epsilon$ $(U, V) \leftarrow (U_t, V_t)$

*Extensions et autres approches*

Le principe d'optimisation alternée d'un critère généralisant la notion d'inertie s'est révélé très fécond en classification, et a connu de nombreux développements. Nous ne mentionnerons ici que quelques extensions utiles en diagnostic.

Une limitation évidente de l'algorithme FCM est la nécessité de définir à l'avance une métrique unique  $A$  pour le calcul des distances aux centres de chacune des classes, ce qui présente un inconvénient lorsque les classes ont des formes ou des volumes différents. Une réponse à ce problème consiste à associer une matrice  $A_k$  à chaque classe  $k$ , et à déterminer les  $A_k$  par apprentissage : on parle alors de distances adaptatives [GOV 75]. Cette approche a été introduite en classification floue par Gustafson et Kessel [GUS 79], dont la méthode consiste à minimiser le critère suivant :

$$J_m^{GK}(U, V, \mathbf{A}) = \sum_{k=1}^M \sum_{i=1}^N u_{i,k}^m \|x_i - \mathbf{v}_k\|_{A_k}^2 \quad [3.21]$$

par rapport à  $U, V$  et à l'ensemble  $\mathbf{A} = \{A_k\}_{k=1}^M$ , sous les contraintes

$$\det(A_k) = \rho_k \quad \forall k \quad [3.22]$$

Gustafson et Kessel ont montré que la minimisation de  $J_m^{GK}$  par rapport à  $\mathbf{A}$  conduisait à la condition d'optimalité supplémentaire suivante :

$$A_k = [\rho_k \det(\tilde{\Sigma}_k)]^{1/n} \tilde{\Sigma}_k^{-1} \quad \forall k \quad [3.23]$$

où  $\tilde{\Sigma}_k$  est la matrice de covariance floue de la class  $k$  définie par

$$\tilde{\Sigma}_k = \left( \sum_{i=1}^N u_{i,k}^m \right)^{-1} \sum_{i=1}^N u_{i,k}^m (\mathbf{x}_i - \mathbf{v}_k)(\mathbf{x}_i - \mathbf{v}_k)^T \quad [3.24]$$

L'algorithme de Gustafson et Kessel (GK) consiste à intégrer le calcul des  $A_k$  par l'équation [3.23] dans l'algorithme FCM, et à remplacer dans [3.19] la matrice  $A$  par  $A_k$  pour le calcul des distances aux centres. Cet algorithme étant plus sensible que FCM aux conditions initiales, une stratégie souvent recommandée consiste à prendre comme solution initiale de GK la partition obtenue par FCM.

Une autre variante intéressante de l'algorithme FCM est l'algorithme PCM (*Possibilistic c-means*) [KRI 93] dont l'objectif est la recherche d'une matrice  $U$  de terme général  $u_{i,k} \in [0, 1]$  ne vérifiant pas nécessairement la condition de normalisation [3.13]. Une telle matrice ne définit donc plus une partition floue de l'ensemble d'apprentissage, mais traduit les « degrés de compatibilité » des vecteurs d'apprentissage avec chacune des classes.

L'algorithme PCM consiste à appliquer le principe de minimisation alternée à un nouveau critère obtenu par addition d'un terme supplémentaire au critère  $J_m$  :

$$J_m^{PCM}(U, V) = \sum_{k=1}^M \sum_{i=1}^N u_{i,k}^m \|x_i - \mathbf{v}_k\|_A^2 + \sum_{k=1}^M w_k \sum_{i=1}^N (1 - u_{i,k})^m \quad [3.25]$$

Les coefficients  $w_k$  sont des nombres positifs fixés. Krishnapuram et Keller ont montré l'existence des conditions nécessaires d'optimalité suivantes pour ce critère :

$$u_{i,k} = \left[ 1 + (\|\mathbf{x}_i - \mathbf{v}_k\|_A^2 / w_k)^{1/(m-1)} \right]^{-1} \quad \forall i, k \quad [3.26]$$

$$\mathbf{v}_k = \frac{\sum_{i=1}^N u_{i,k}^m \mathbf{x}_i}{\sum_{i=1}^N u_{i,k}^m} \quad \forall k. \quad [3.27]$$

L'algorithme d'optimisation s'obtient à partir de FCM en remplaçant les équations [3.19] et [3.20] par [3.26] et [3.27]. Krishnapuram et Keller [KRI 93] recommandent d'utiliser FCM pour fournir une solution initiale à l'algorithme PCM, et proposent différentes heuristiques pour le choix des coefficients  $w_k$ .

Notons enfin la similitude entre les algorithmes de classification flous de type FCM et les approches probabilistes basées sur l'estimation de paramètres de mélanges de lois à l'aide de l'algorithme EM [CEL 89], [CEL 95], qui fournissent également une matrice de partition floue sous la forme des probabilités estimées d'appartenance des observations aux classes. Bien que l'algorithme EM ait une justification claire comme méthode de maximisation de la vraisemblance en

présence d'observations incomplètes, rien n'empêche *a priori* l'utilisation de cette technique comme méthode de calcul d'une partition floue, même en l'absence d'hypothèses probabilistes. Il existe à notre connaissance peu de résultats relatifs à la comparaison des approches FCM et EM en classification floue.

L'un des problèmes fondamentaux en classification est le choix du nombre de classes et, plus généralement, la validation d'une partition. On trouvera dans la référence [BEZ 98] une synthèse sur les principaux critères d'évaluation d'une partition floue.

### 3.2.3. Définition des fonctions d'appartenance

#### *Exploitation des résultats de la classification floue*

L'apprentissage ayant été réalisé à partir de données non étiquetées par l'un des algorithmes de classification floue précédents, il convient en général d'interpréter chacune des classes en examinant s'il est possible de lui associer un mode de fonctionnement du système surveillé. Dans le cas où à chaque mode de fonctionnement correspond une seule classe, on pourra déterminer la valeur en tout nouveau point  $\mathbf{x}$  des fonctions d'appartenance par une équation telle que [3.19] ou [3.26] selon l'algorithme de classification utilisé pour l'apprentissage des prototypes. Notons que l'utilisation de [3.19] entraîne

$$\sum_{k=1}^M \mu_k(\mathbf{x}) = 1 \quad \forall \mathbf{x} \quad [3.28]$$

Cela semble une propriété trop forte en diagnostic, notamment lorsque l'ensemble des modes de fonctionnement répertoriés n'est pas exhaustif : il se peut alors que le système se trouve à un instant donné éloigné de tous les modes connus, ce qui devrait se traduire par une condition du type :

$$\max_{k=1, M} \mu_k(\mathbf{x}) \approx 0 \quad [3.29]$$

Il est dans ce cas préférable d'adopter la formule « possibiliste » pour le calcul des degrés d'appartenance

$$\mu_k(\mathbf{x}) = \left[ 1 + (\|\mathbf{x} - \mathbf{v}_k\|_A^2 / w_k)^{1/(m-1)} \right]^{-1} \quad \forall k \quad [3.30]$$

Lorsque plusieurs classes correspondent à un même mode de fonctionnement, il paraît naturel de définir le degré d'appartenance au mode  $\omega_k$  en fonction des degrés d'appartenance aux classes correspondant à ce mode, en

utilisant un opérateur de combinaison disjonctif. Soient  $\mathbf{v}_{k,1}, \dots, \mathbf{v}_{k,n_k}$  les prototypes caractérisant le mode  $k$ , et  $\mu_{k,1}, \dots, \mu_{k,n_k}$  les fonctions d'appartenance correspondantes. On pourra poser, pour une  $t$ -conorme  $\perp$

$$\mu_k(\mathbf{x}) = \mu_{k,1} \perp \dots \perp \mu_{k,n_k} \quad [3.31]$$

Par exemple, Masson [MAS 96] propose d'utiliser l'opérateur de somme bornée :

$$\mu_k(\mathbf{x}) = \min \left[ 1, \sum_{\ell=1}^{n_k} \mu_{k,\ell} \right] \quad [3.32]$$

L'emploi d'autres opérateurs tels que le maximum pourrait se justifier également.

#### *Apprentissage supervisé des fonctions d'appartenance*

Dans certains cas, on dispose d'un ensemble d'apprentissage *étiqueté* de la forme  $\mathcal{E} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ ,  $y_i$  étant l'étiquette floue de l'exemple  $i$  :

$$y_i = (\mu_{i,1}, \dots, \mu_{i,M}) \quad [3.33]$$

Dans l'expression ci-dessus,  $\mu_{i,k}$  désigne le degré d'appartenance du système au mode de fonctionnement  $k$ , pour l'exemple  $i$ . Ces degrés d'appartenance peuvent avoir été obtenus par interrogation d'un expert, ou par l'un des algorithmes de classification floue décrits au paragraphe précédent.

Le problème consiste dans ce cas à interpoler les données d'apprentissage de manière à fournir une évaluation des fonctions d'appartenance pour tout nouveau point  $\mathbf{x}$  [BEL 66]. Parmi les méthodes qui ont été proposées pour résoudre ce problème, on peut citer les extensions floues de la méthode des  $K$  plus proches voisins, et les réseaux de neurones (cf. Chapitre 4).

Des extensions floues de la méthode des  $K$  plus proches voisins ont été proposées par Joswik [JOZ 83], Keller *et al.* [KEL 85], Bezdek *et al.* [BEZ 86] et Béreau [BÉR 86]. L'approche décrite par Joswik et Bezdek consiste à définir le degré d'appartenance de  $\mathbf{x}$  à la classe  $k$  comme la moyenne des degrés d'appartenance à cette classe parmi les  $K$  plus proches voisins de  $\mathbf{x}$  :

$$\mu_k(\mathbf{x}) = \frac{1}{K} \sum_{i \in \Phi_K(\mathbf{x})} \mu_{i,k} \quad [3.34]$$

où  $\Phi_K(\mathbf{x})$  désigne l'ensemble des indices des  $K$  plus proches voisins de  $\mathbf{x}$  dans l'ensemble d'apprentissage, selon une certaine métrique (par exemple, la métrique euclidienne). Les règles proposées par Keller et Béreau sont *a priori* plus

satisfaisantes car elles font intervenir également les distances aux voisins de  $\mathbf{x}$ . Keller [KEL 85] propose l'expression suivante :

$$\mu_k(\mathbf{x}) = \frac{\sum_{i \in \Phi_K(\mathbf{x})} u_{i,k} \|\mathbf{x} - \mathbf{x}_i\|^{-2/(m-1)}}{\sum_{i \in \Phi_K(\mathbf{x})} \|\mathbf{x} - \mathbf{x}_i\|^{-2/(m-1)}} \quad [3.35]$$

où  $m$  est un coefficient. Notons que, si les  $u_{i,k}$  vérifient la contrainte de normalisation [3.13], alors la condition [3.28] est nécessairement vérifiée, ce qui, comme nous l'avons déjà souligné, n'est pas toujours souhaitable en diagnostic. La règle proposée par Béreau [BÉR 86] consiste à poser :

$$\mu_k(\mathbf{x}) = \max_{i \in \Phi_k(\mathbf{x})} u_{i,k} \exp \left[ -\eta \left( \frac{\|\mathbf{x} - \mathbf{x}_i\|}{d_k} \right)^2 \right] \quad [3.36]$$

où  $d_k$  est la distance moyenne entre vecteurs d'apprentissage ayant un degré d'appartenance à la classe  $\omega_k$  supérieur à un certain seuil, et  $\eta$  un coefficient. Par rapport à la précédente, cette règle présente la particularité d'affecter des degrés d'appartenance faibles aux points éloignés de l'ensemble d'apprentissage.

Une autre approche pour l'apprentissage des fonctions d'appartenance par interpolation est basée sur l'utilisation de réseaux de neurones. Par exemple, Millemann et Lengellé [MIL 95] proposent d'approcher les évaluations subjectives de degrés d'appartenance par les sorties d'un perceptron multicouches, en minimisant le « degré de flou » des sorties. L'utilisation de réseaux à fonctions de base radiales pour l'évaluation de degrés d'appartenance a par ailleurs été étudiée par Masson [MAS 92] et Peltier [PEL 93a]. L'intérêt de ce type de réseau en diagnostic est qu'il garantit l'obtention de degrés d'appartenance proches de 0 pour des entrées atypiques.

#### 3.2.4. *Décision à partir des degrés d'appartenance*

Ayant calculé le vecteur  $\mathbf{z}$  des degrés d'appartenance aux différentes classes

$$\mathbf{z} = \phi(\mathbf{x}) = (\mu_1(\mathbf{x}), \dots, \mu_M(\mathbf{x}))^T \quad [3.37]$$

il reste à en déduire le choix d'une action  $\alpha = \delta(\mathbf{z}) \in \mathcal{A}$ . Pour cela, il est nécessaire de s'appuyer sur des principes de décision ayant pour base les degrés d'appartenance. Une étude comparative de différentes règles a été menée par Masson *et al.* [MAS 96]. Nous reprendrons ici les grandes lignes de cette étude.

Dans le cas le plus simple où les seules actions autorisées sont l'affectation aux classes :  $\mathcal{A} = \{\alpha_1, \dots, \alpha_M\}$ ,  $\alpha_k$  désignant l'affectation à la classe  $\omega_k$ , il est naturel d'appliquer un principe de *maximum d'appartenance* [PAL 77],

consistant à choisir la classe à laquelle le vecteur  $\mathbf{x}$  a le plus haut degré d'appartenance :

$$D(\mathbf{x}) = \alpha_k \Leftrightarrow \mu_k(\mathbf{x}) > \mu_\ell(\mathbf{x}) \quad \forall \ell \neq k \quad [3.38]$$

Plusieurs mécanismes peuvent être envisagés pour le cas plus général où l'ensemble  $\mathcal{A}$  des actions inclut, outre l'affectation aux classes, les options de rejet en distance  $\alpha_d$  et en ambiguïté  $\alpha_a$ . Le principe le plus simple est celui de la règle dite « avec seuillage des appartenances », qui consiste à associer à chaque classe  $k$  un seuil d'appartenance  $s_k$  défini, par exemple, de façon à ce qu'une proportion donnée de vecteurs d'apprentissage aient un degré d'appartenance à la classe  $k$  supérieur à  $s_k$ . Soit, pour un vecteur  $\mathbf{x}$  donné, l'ensemble  $J(\mathbf{x})$  des classes  $k$  auxquelles  $\mathbf{x}$  a un degré d'appartenance supérieur à  $s_k$  :

$$J(\mathbf{x}) = \{k \in \{1, \dots, M\} \mid \mu_k(\mathbf{x}) > s_k\} \quad [3.39]$$

Une règle simple consiste à rejeter  $\mathbf{x}$  en ambiguïté si  $\mathbf{x}$  a un degré d'appartenance significatif à plusieurs classes, et à choisir l'option de rejet de distance si  $J(\mathbf{x})$  est vide :

$$D(\mathbf{x}) = \begin{cases} \alpha_d & \text{si } J(\mathbf{x}) = \emptyset \\ \alpha_a & \text{si } |J(\mathbf{x})| > 1 \\ \alpha_k & \text{si } J(\mathbf{x}) = \{k\} \quad k \in \{1, \dots, M\} \end{cases} \quad [3.40]$$

L'inconvénient de la règle précédente est le fait que les deux options de rejet sont contrôlées par les mêmes paramètres  $s_k$ . La règle dite « du rapport d'appartenance », proposée par Frélicot [FRÉ 92], [FRÉ 93] s'appuie sur le rapport :

$$R = \frac{\mu_m(\mathbf{x})}{\mu_p(\mathbf{x})} \quad [3.41]$$

avec :

$$\begin{aligned} \mu_p(\mathbf{x}) &= \max_{k \in J(\mathbf{x})} \mu_k(\mathbf{x}) \\ \mu_m(\mathbf{x}) &= \max_{k \in J(\mathbf{x}) \setminus \{p\}} \mu_k(\mathbf{x}) \end{aligned}$$

On a évidemment  $0 \leq R \leq 1$ , et il y a ambiguïté quand  $R$  est proche de 1. On pose alors la règle suivante :

$$D(\mathbf{x}) = \begin{cases} \alpha_d & \text{si } J(\mathbf{x}) = \emptyset \\ \alpha_a & \text{si } |J(\mathbf{x})| > 1 \text{ et } R \geq T \\ \alpha_p & \text{sinon} \end{cases} \quad [3.42]$$

$T$  est un paramètre fixé par l'utilisateur, permettant le contrôle du rejet en ambiguïté.



Enfin, une troisième règle de décision avec rejet, dit « règle des degrés de confiance », a été proposée par Masson *et al.* [MAS 96], [FRÉ 95]. Cette règle s'appuie sur la définition de degrés de confiance fonctions des  $M$  degrés d'appartenance. Plus précisément, on définit le degré de confiance associé à la classe  $k$  pour le vecteur forme  $\mathbf{x}$  comme

$$c_k(\mathbf{x}) = \mu_k(\mathbf{x}) \prod_{\ell \neq k} [1 - \mu_\ell(\mathbf{x})]. \quad [3.43]$$

La quantité  $c_k(\mathbf{x})$  est d'autant plus grande que le degré d'appartenance à la classe  $k$  est élevé, et que les degrés d'appartenance aux autres classes sont faibles. Par ailleurs, soit

$$c_0(\mathbf{x}) = \prod_{\ell=1}^M [1 - \mu_\ell(\mathbf{x})] \quad [3.44]$$

Cette quantité est d'autant plus grande que les degrés d'appartenance aux classes sont faibles, et traduit donc le caractère « atypique » du vecteur  $\mathbf{x}$ . Une règle de décision simple basée sur les degrés de confiance est donc la suivante. Soit  $q$  l'indice tel que

$$c_q(\mathbf{x}) = \max_{k=0, M} c_k(\mathbf{x}) \quad [3.45]$$

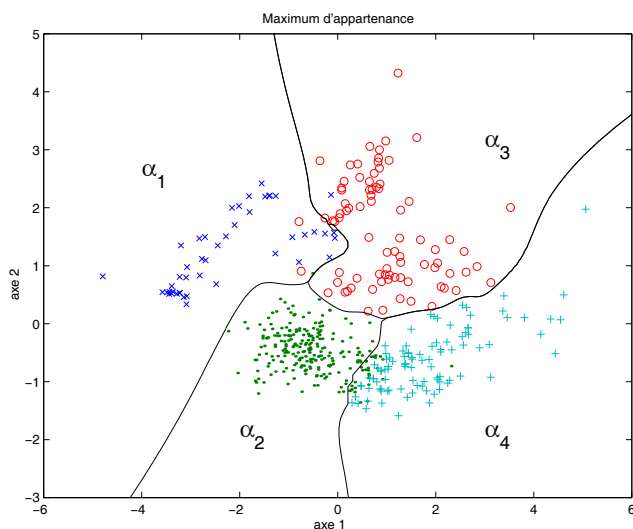
On pose

$$D(\mathbf{x}) = \begin{cases} \alpha_d & \text{si } q = 0 \text{ et } c_0(\mathbf{x}) > \kappa \\ \alpha_k & \text{si } q = k \text{ et } c_k(\mathbf{x}) > \kappa \\ \alpha_a & \text{sinon,} \end{cases} \quad [3.46]$$

$\kappa$  étant un coefficient contrôlant à la fois le rejet d'ambiguïté et le rejet de distance.

L'étude de ces différentes règles, menée par Masson *et al.* [MAS 96], a mis en évidence des comportements similaires pour les règles du rapport d'appartenance et des degrés de confiance, cette dernière présentant l'avantage de ne nécessiter le réglage que d'un seuil.

EXEMPLE : La démarche de construction d'un système flou de reconnaissance de formes présentée ci-dessus a été appliquée à un ensemble d'apprentissage issu d'une application en surveillance de l'environnement [TRA 95] (le même exemple est utilisé dans le chapitre 4 de ce volume). Il s'agit de mesures de quatre paramètres physico-chimiques (pH, conductivité, concentrations en ammoniaque et en nitrites) effectuées quotidiennement pendant deux ans sur l'eau de Seine en amont d'une usine de traitement d'eau potable. Une étude préliminaire a conduit à la définition de quatre classes de qualité d'eau. La dimension de l'espace de représentation a été ramenée à deux par analyse en composantes principales.

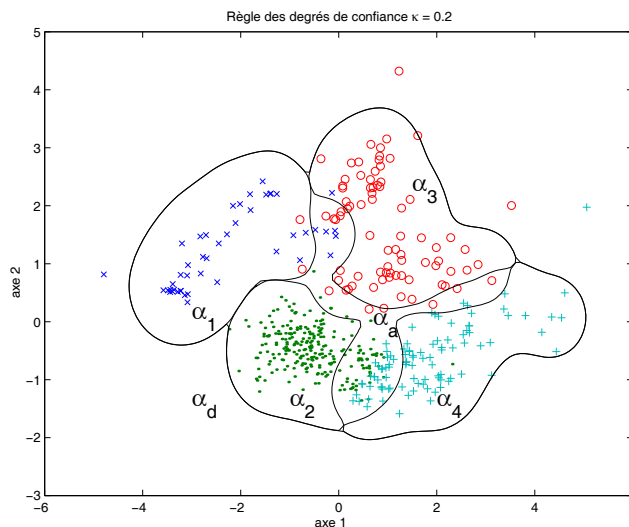


**Figure 3.2.** Données de qualité d'eau : régions de décision pour la règle du maximum d'appartenance.

Pour chaque classe, dix prototypes ont été générés par l'algorithme FCM, et la fonction d'appartenance a été définie par les équations [3.30] (avec  $m = 1.5$  et  $w_k = 0.3, \forall k$ ) et [3.32]. Les régions de décision correspondant aux règles du maximum d'appartenance et des degrés de confiance sont illustrées sur les figures 3.2 et 3.3. On constate que cette seconde règle permet bien le rejet en ambiguïté des points situés entre les classes, et le rejet en distance des points atypiques.

### 3.2.5. Compléments bibliographiques

Nous nous sommes volontairement limités dans cette exposé à une présentation des principales méthodes de référence en reconnaissance de formes floues pour le diagnostic. Plusieurs approches floues en reconnaissance des formes, moins courantes en diagnostic, ont été omises, parmi lesquelles on peut citer les systèmes flous [KAS 96], les arbres de décision flous [GIO 98], les systèmes neuro-flous [NAU 98], et la technique d'appariement flou (*fuzzy pattern matching*) [DEV 98], [DEV 99]. La référence [BEZ 92] contient un ensemble d'articles importants dans le domaine parus avant 1992. Des ouvrages de référence récents tels que [RUS 98a] et [BEZ 99], proposent un panorama assez complet des applications de la théorie des ensembles flous en reconnaissance des formes.



**Figure 3.3.** Données de qualité d'eau : régions de décision pour la règle des degrés de confiance.

L'adaptation d'un système de décision en environnement non stationnaire (évolution des classes, apparition de nouveaux modes de fonctionnement) est un problème fondamental en diagnostic, mais non spécifiquement lié au choix d'un formalisme particulier de représentation de l'incertitude. On trouvera des développements très complets sur ces questions dans les références [PEL 93c], [PEL 93a], [ZIE 94], [ZIE 95], [BOU 96a], [BOU 97], [DEV 99], entre autres.

Les techniques de reconnaissance de formes floues ont donné lieu à de nombreuses applications dans des domaines très variés, parmi lesquels : la détection d'hypovigilance d'un conducteur automobile [PEL 93a], [PEL 93b], le suivi de la qualité d'usinage [ZIE 94], [ZIE 95], la gestion en temps réel du réseau téléphonique français [BOU 93], [BOU 96b], [CRO 99], le contrôle non destructif [DEV 99] et la surveillance de l'environnement [BOU 99].

### 3.3. Approche crédibiliste

#### 3.3.1. Cadre général

Comme nous l'avons montré dans le paragraphe 3.2 de ce chapitre, la théorie des sous-ensemble flous se révèle être un outil bien adapté à la modélisation de situations dans lesquelles, la définition des modes de fonctionnement d'un

système étant imprécise, les transitions d'un mode à un autre peuvent être considérées comme graduelles; la variable  $y$  décrivant l'état du système à un instant donné peut alors être considérée comme prenant simultanément, à des degrés divers, plusieurs valeurs dans  $\Omega$ .

Nous revenons dans ce paragraphe à une formalisation plus « classique » des problèmes de diagnostic dans laquelle le système, sur lequel est effectuée une mesure  $\mathbf{x}$ , est supposé se trouver à un instant donné dans un et un seul mode de fonctionnement  $y \in \Omega$ . Le problème considéré consiste à caractériser la connaissance partielle apportée sur  $y$  par un ensemble d'apprentissage composé de  $N$  observations, *éventuellement imprécises ou incertaines*, des variables  $\mathbf{x}$  et  $y$ .

La problématique de l'apprentissage à partir de données partiellement ou imparfaitement étiquetées a, jusqu'à présent, fait l'objet d'assez peu de travaux en reconnaissance de formes. Or, ce type de données n'est pas rare en diagnostic, du fait soit de l'évolution des connaissances conduisant à affiner la définition des classes, soit de la difficulté à reconnaître, même a posteriori, l'état d'un système. Par exemple, Richard [RIC 98], [RIC 99] décrit un problème de reconnaissance de complexes K dans des signaux électroencéphalographiques, pour lequel des désaccords importants sont constatés entre experts chargés de l'étiquetage des signaux [DEN 99a].

Par ailleurs, il importe dans de nombreuses applications de prendre en compte l'incertitude non seulement sur les entrées du système d'apprentissage, mais également sur les sorties. Dans de nombreux problèmes de diagnostic, les prédictions fournies par une méthode statistique ne sont que des éléments parmi d'autres permettant à un expert de prendre une décision. Il est souhaitable dans ce cas que l'incertitude sur chacune des prédictions soit décrite de manière suffisamment fidèle. De même, dans les problèmes de fusion de décisions, la stratégie de combinaison des informations doit pouvoir prendre en compte la plus ou moins grande fiabilité des experts ou capteurs fusionnés.

Les considérations précédentes nous ont conduit à proposer une formulation assez générale du problème de l'apprentissage statistique supervisé, dans laquelle les entrées et les sorties du système sont représentées à l'aide du formalisme des fonctions de croyance [SHA 76], [SME 94]. Nous adoptons pour cela le modèle des croyances transférables (MCT) de Smets, dont nous rappelons brièvement ci-dessous les principaux éléments. Plusieurs extensions permettant la représentation d'informations floues seront également mentionnées.

### 3.3.2. Théorie des fonctions de croyance

#### Concepts fondamentaux

La théorie des fonctions de croyance est issue des travaux de Dempster sur les bornes inférieure et supérieure d'une famille de distributions de probabilités [DEM 67]. Elle a été développée par Shafer [SHA 76] qui a montré l'intérêt des fonctions de croyance comme formalisme de représentation de l'incertitude. Une interprétation subjectiviste et normative en a été donnée sous le nom Modèle des Croyances Transférables (MCT) par Smets, qui a posé des bases axiomatiques et clarifié les liens entre représentation des croyances et prise de décision [SME 94], [SME 98b]; nous adopterons ce point de vue dans la suite de ce chapitre. Nous nous contenterons tout d'abord de rappeler les principaux éléments du MCT, afin notamment de préciser les notations et la terminologie utilisées par la suite.

De manière très générale, le problème posé consiste à modéliser la croyance d'un agent rationnel relativement à la valeur prise par une variable  $y$ , sur la base d'un corpus de connaissance. On suppose que cette croyance peut être représentée par l'allocation d'une « masse de croyance » égale à un, à certains sous-ensembles du domaine  $\Omega$  de  $y$  (supposé fini), appelés *éléments focaux*. On appelle *structure de croyance* (SC) l'application qui, à chaque sous-ensemble  $A$  de  $\Omega$ , associe la fraction de masse  $m(A)$  affectée à  $A$ . On a donc nécessairement :

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad [3.47]$$

Remarquons que l'on n'impose pas systématiquement la condition  $m(\emptyset) = 0$ , la masse allouée à  $\emptyset$  pouvant recevoir une interprétation dans le cadre de l'hypothèse du monde ouvert [SME 90].

A la SC  $m$  est associée une *fonction de croyance*  $\text{bel}_m$ , qui associe à toute partie  $A$  de  $\Omega$  la somme des masses allouées aux sous-ensembles de  $A$  :

$$\text{bel}_m(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad [3.48]$$

Chaque nombre  $\text{bel}_m(A) \in [0, 1]$  s'interprète comme un *degré de croyance* en l'hypothèse «  $y \in A$  ». On montre que la fonction  $\text{bel}_m$  est une capacité de Choquet d'ordre infini, et qu'il existe une bijection entre l'ensemble des SC et celui des fonctions de croyance [SHA 76]. Smets [SME 97] a montré que l'usage de telles fonctions pour la représentation des croyances pouvait être déduit d'une série d'axiomes.

Supposons maintenant que deux sources d'informations distinctes induisent séparément deux SC  $m_1$  et  $m_2$ . La considération simultanée de ces deux sources

induit une nouvelle SC,  $m_1 \cap m_2$ , appelée somme conjonctive de  $m_1$  et  $m_2$ , et définie par :

$$(m_1 \cap m_2)(A) = \sum_{B \cap C = A} m_1(B)m_2(C) \quad \forall A \subseteq \Omega \quad [3.49]$$

On montre facilement l'associativité et la commutativité de l'opérateur  $\cap$ . Notons que l'on peut définir, à partir d'autres opérations ensemblistes, d'autres opérateurs sur les SC comme, par exemple, la somme disjonctive [SME 93], qui correspond à un mode de combinaison beaucoup plus « prudent » (on suppose que l'une *ou* l'autre des deux sources doit être prise en considération).

Un procédé commode permettant de limiter l'influence d'une source d'information dans une somme conjonctive est l'opération d'*affaiblissement*, qui consiste à multiplier chacune des masses de croyance affectées aux sous-ensembles stricts de  $\Omega$  par un degré de confiance  $(1 - \beta)$  en la fiabilité de la source [SHA 76]. La SC  $m^\beta$  obtenue à partir de  $m$  par affaiblissement de facteur  $\beta$  est définie par :

$$m^\beta(A) = (1 - \beta)m(A) \quad \forall A \subseteq \Omega, A \neq \Omega \quad [3.50]$$

La masse soustraite aux sous-ensembles stricts de  $\Omega$  est par ailleurs reportée sur  $\Omega$ . Shafer [SHA 76] souligne l'intérêt de cette opération pour prendre en compte la fiabilité des observations dans les applications statistiques des fonctions de croyance.

Enfin, le MCT opère, contrairement au modèle Bayésien, une distinction très nette entre la représentation des croyances et la prise de décision, cette dernière devant être basée sur le formalisme des probabilités afin de garantir la cohérence des décisions [SME 94]. On appelle *fonction de probabilité pignistique* associée à une SC  $m$  la fonction  $P_m : \Omega \rightarrow [0, 1]$  définie par :

$$P_m(\omega) = \sum_{B \subseteq \Omega} \frac{m^*(B)}{|B|} \mu_B(\omega) \quad [3.51]$$

$m^*$  étant la SC normalisée associée à  $m$ , et  $\mu_B$  la fonction indicatrice de l'ensemble  $B$ . Face à un ensemble d'actions possibles, le MCT préconise l'action de risque minimal au sens de la probabilité pignistique.

### 3.3.3. Construction des fonctions de croyance en discrimination

#### *Principe*

Soit un échantillon de  $N$  individus extraits d'une population  $\mathcal{P}$ , et décrits par deux variables  $\mathbf{x}$  et  $y$ , à valeurs respectivement dans  $\mathfrak{R}^d$  et  $\Omega$ . On note  $\mathbf{x}_i$  et

$y_i$  les valeurs prises par les variables  $\mathbf{x}$  et  $y$  pour l'individu  $i$ . Pour chaque  $i$ , la variable  $\mathbf{x}_i$  est supposée connue, mais la variable  $y_i$  est partiellement inconnue, et décrite par une SC  $m_i$  sur  $\Omega$ . On dispose donc d'un ensemble d'apprentissage de la forme :

$$\mathcal{E} = \{e_i = (\mathbf{x}_i, m_i)\}_{i=1}^N \quad [3.52]$$

Il convient ici de remarquer que cette formulation très générale englobe un grand nombre de situations distinctes, dont certaines ont déjà été étudiées dans un cadre plus restrictif [DEN 99a]. En particulier, dans le cas où chaque SC  $m_i$  a un seul élément focal singleton, il s'agit d'un problème classique d'apprentissage supervisé avec étiquetage précis et certain. Lorsque chaque SC  $m_i$  a un seul élément focal  $F_i \subseteq \Omega$ , l'étiquetage est imprécis : on ignore la valeur exacte de  $y_i$ , mais on sait avec certitude que  $y_i \in F_i$ . Le cas le plus général où  $m_i$  est une SC quelconque possédant un nombre arbitraire d'éléments focaux (l'étiquetage étant alors à la fois imprécis et incertain) n'a pas encore, à notre connaissance, été étudié, bien qu'il soit adapté à la modélisation de nombreux problèmes réels [DEN 99a].

Etant donné un ensemble d'apprentissage de la forme décrite par l'équation [3.52], on considère maintenant un nouvel individu sur lequel on observe la variable  $\mathbf{x}$ , et pour lequel on cherche à prédire la variable  $y$ . Une telle prédiction étant nécessairement incertaine, on cherche à décrire l'information relative à  $y$  sous la forme d'une SC  $m_y$  sur  $\Omega$  (la notation  $m_y$  étant introduite pour mettre nettement en évidence la variable à laquelle se rapporte la SC). La fonction  $\phi$  décrivant l'incertitude sur la connaissance de l'état du système est dans ce cas une fonction de  $\mathfrak{R}^d$  dans l'ensemble  $\mathcal{S}(\Omega)$  des SC sur  $\Omega$ .

La méthode proposée dans [DEN 94], [DEN 95b] pour construire une telle fonction repose sur deux principes fondamentaux : l'affaiblissement des SC  $m_i$  en fonction d'une mesure de distance (ou, plus généralement, de dissimilarité)  $\Delta$  sur  $\mathfrak{R}^d$ , et la combinaison des SC résultantes par l'opérateur de somme conjonctive.

Considérons tout d'abord l'impact d'un exemple d'apprentissage  $e_i = (\mathbf{x}_i, m_i)$  sur notre croyance relativement à la valeur de  $y$ . Afin d'introduire la méthode dans un cas simple, supposons tout d'abord que  $m_i$  ait un seul élément focal  $F_i$ , et que  $\mathbf{x} = \mathbf{x}_i$ . Il est donc certain que  $y_i$  appartient à  $F_i$ . Ce fait accrédite l'hypothèse selon laquelle  $y$  appartient également à  $F_i$ , et aucune autre hypothèse. S'il n'existe pas *a priori* de relation fonctionnelle entre  $\mathbf{x}$  et  $y$ , on n'aura cependant aucune certitude relativement à la valeur de  $y$ , ce qui conduit à poser :

$$m_y(F_i|e_i) = \alpha \quad [3.53]$$

$$m_y(\Omega|e_i) = 1 - \alpha, \quad 0 < \alpha \leq 1 \quad [3.54]$$

Si maintenant  $\mathbf{x} \neq \mathbf{x}_i$ , il est naturel de faire dépendre la masse de croyance  $m_y(F_i|e_i)$  de la *distance* entre  $\mathbf{x}$  et  $\mathbf{x}_i$  en posant :

$$m_y(F_i|e_i) = \varphi[\Delta(\mathbf{x}, \mathbf{x}_i)] \quad [3.55]$$

$$m_y(\Omega|e_i) = 1 - \varphi[\Delta(\mathbf{x}, \mathbf{x}_i)] \quad [3.56]$$

$\varphi$  est une fonction décroissante vérifiant

$$\varphi(0) \leq 1 \quad [3.57]$$

et

$$\lim_{\Delta \rightarrow \infty} \varphi(\Delta) = 0 \quad [3.58]$$

La généralisation au cas où la SC  $m_i$  est de forme quelconque est immédiate si l'on remarque que  $m_y$  définie par les équations [3.55] et [3.56] s'obtient à partir de  $m_i$  par l'opération d'affaiblissement, le facteur d'affaiblissement étant égal à  $1 - \varphi[\Delta(\mathbf{x}, \mathbf{x}_i)]$ . En étendant ce principe, on obtient donc dans le cas général :

$$m_y(F|e_i) = \varphi[\Delta(\mathbf{x}, \mathbf{x}_i)]m_i(F) \quad \text{si } F \neq \Omega \quad [3.59]$$

$$m_y(\Omega|e_i) = 1 - \varphi[\Delta(\mathbf{x}, \mathbf{x}_i)] \sum_{\{G \neq \Omega\}} m_i(G). \quad [3.60]$$

Ayant quantifié sous la forme d'une SC  $m_y(\cdot|e_i)$  l'impact de chaque exemple d'apprentissage  $e_i$  sur notre croyance en la valeur de  $y$ , il nous reste maintenant à combiner les  $N$  SC induites par les  $N$  exemples  $e_i$ ,  $i = 1, \dots, N$ . Plusieurs opérateurs de combinaison peuvent *a priori* être envisagés, mais il existe selon nous un argument décisif en faveur de la somme conjonctive. Supposons que, pour un exemple  $e_i$ , on n'ait aucune information sur  $y$  :  $m_i$  est donc la SC vide et l'on a  $m_i(\Omega) = 1$ , d'où également  $m_y(\Omega|e_i) = 1$ . Or, un tel exemple n'apportant aucune information sur  $y$ , la SC vide  $m_y(\cdot|e_i)$  doit être un élément neutre de l'opérateur de combinaison, condition vérifiée pour la somme conjonctive et non pour la somme disjonctive ou la moyenne. Nous définirons donc la SC  $m$  induite par la totalité de l'ensemble d'apprentissage  $\mathcal{E}$  comme la somme conjonctive des SC induites par chacun des éléments de  $\mathcal{E}$  :

$$m_y(\cdot|\mathcal{E}) = \bigcap_{i=1}^N m_y(\cdot|e_i) \quad [3.61]$$

La fonction  $m_y(\cdot|\mathcal{E})$  modélise donc l'information apportée par l'ensemble d'apprentissage, relativement à la variable à expliquer  $y$ . Plusieurs stratégies de décision en discrimination basées sur cette information (affectation à une ou plusieurs classes, rejet, etc.) ont été étudiées dans [DEN 97a], [DEN 98a] et seront évoquées au paragraphe 3.3.4. Remarquons au passage que le même



principe est applicable en régression, une prédiction ponctuelle de  $y$  pouvant être obtenue en calculant l'espérance de  $y$  relativement à la mesure de probabilité pignistique [PET 99b], [PET 99c], [PET 99a]. Enfin, signalons l'existence d'autres approches pour la construction de fonctions de croyance à partir de données, basées par exemple sur la fonction de vraisemblance [APP 98] ou sur l'induction d'arbres de décision [ELO 00], [DEN 00d].

### *Optimisation des performances*

Nous avons jusqu'à présent passé sous silence le problème du choix de la mesure de dissimilarité  $\Delta$  et de la fonction  $\varphi$  dans les équations [3.55]-[3.60]. Or, il s'agit de toute évidence d'un problème fondamental puisque ces choix conditionneront les performances de la méthode. Une approche pour résoudre ce problème consiste à choisir  $\Delta$  et  $\varphi$  parmi une famille paramétrée de fonctions. Notons

$$\Phi_{\theta}(\mathbf{x}, \mathbf{x}_i) = \varphi_{\theta_2}[\Delta_{\theta_1}(\mathbf{x}, \mathbf{x}_i)] \quad [3.62]$$

où  $\theta_1$  est un paramètre définissant la distance,  $\theta_2$  un paramètre définissant la fonction  $\varphi$ , et  $\theta = (\theta_1, \theta_2)^T$ . Un choix naturel dans le cas, considéré jusqu'à présent, où les composantes du vecteur forme  $\mathbf{x}$  sont des variables numériques, consiste à choisir  $\Delta$  dans la famille des distances euclidiennes généralisées :

$$\Delta_{\theta_1}(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} - \mathbf{x}_i)^T A (\mathbf{x} - \mathbf{x}_i)]^{1/2} \quad [3.63]$$

$A$  étant une matrice symétrique définie positive. On peut dans ce cas justifier [DEN 98a] le choix suivant pour  $\varphi$  :

$$\varphi_{\theta_2}(\Delta) = \alpha \exp(-\gamma \Delta^2) \quad [3.64]$$

avec

$$\alpha \in ]0, 1[, \quad [3.65]$$

$$\gamma \in ]0, +\infty[, \quad [3.66]$$

$$\theta_2 = (\alpha, \gamma)^T \quad [3.67]$$

Une fois le choix de la famille  $\{\Phi_{\theta} | \theta \in \Theta\}$  effectué, il est possible d'optimiser les performances du système d'apprentissage par une technique de validation croisée [ZOU 95], [ZOU 97a], [ZOU 98]. Soit  $\mathcal{E}_{-i}$  l'ensemble d'apprentissage privé de l'exemple  $i$ , et  $m_{-i} = m_{y_i}(\cdot | \mathcal{E}_{-i})$  la SC représentant l'information apportée par  $\mathcal{E}_{-i}$  sur  $y_i$ , obtenue en appliquant les équations [3.59]-[3.61] à  $\mathbf{x}_i$ . Etant donné un critère d'erreur  $\rho(m_{-i}, m_i)$ , il s'agit de trouver la valeur  $\hat{\theta}$  du paramètre  $\theta$  minimisant l'erreur moyenne sur l'ensemble d'apprentissage :

$$\hat{\theta} = \arg \min_{\theta} E(\theta) \quad [3.68]$$

$$E(\theta) = \sum_{i=1}^N \rho(m_{-i}, m_i). \quad [3.69]$$

Plusieurs choix de critères d'erreur ont été proposés pour la discrimination [ZOU 98], et la régression [PET 99b], [PET 99c], [PET 99a].

#### *Réduction du temps de calcul*

Dans le cas où les SC  $m_i$  ont une forme générale et où  $N$  est grand, la combinaison des  $N$  SC  $m_y(\cdot|e_i)$  peut devenir pénalisante par suite de la complexité exponentielle de la somme conjonctive. Plusieurs approches peuvent être envisagées pour remédier à ce problème.

Une première méthode, initialement proposée dans [DEN 95b], consiste à n'utiliser que les  $K$  plus proches voisins de  $\mathbf{x}$  dans  $\mathcal{E}$  (ou, alternativement, les voisins situés dans une boule de rayon  $r$  centrée sur  $\mathbf{x}$ ); néanmoins, cette approche nécessite encore le calcul de toutes les distances entre  $\mathbf{x}$  et les  $\mathbf{x}_i$ .

Une autre méthode consiste à résumer l'ensemble d'apprentissage sous la forme d'un nombre restreint de prototypes [DEN 95a], [DEN 97b], [DEN 00c], [PET 99c], éventuellement optimisés par minimisation d'un critère d'erreur.

#### **3.3.4. Principes de décision**

Supposons maintenant que l'incertitude relative à l'état du système décrit par un vecteur forme  $\mathbf{x}$  soit représentée par une structure de croyance  $m_y(\cdot|\mathcal{E})$  définie sur l'ensemble des classes  $\Omega$ , et considérons le problème du choix d'une action parmi un ensemble fini  $\mathcal{A}$  [DEN 97a]. La mise en œuvre d'une action  $\alpha \in \mathcal{A}$  lorsque l'individu appartient en réalité à la classe  $\omega \in \Omega$  est supposée entraîner un coût noté  $C(\alpha, \omega)$ . Le risque pignistique associé au choix de  $\alpha$  se définit comme l'espérance du coût, relativement à la distribution de probabilité pignistique [3.51]. En notant  $p_y$  la fonction de probabilité pignistique associée à  $m_y(\cdot|\mathcal{E})$ , on a donc :

$$R(\alpha) = \sum_{k=1}^M C(\alpha, \omega_k) p_y(\omega_k) \quad [3.70]$$

$$= \sum_{k=1}^M C(\alpha, \omega_k) \sum_{B \ni \omega_k} \frac{m_y(B|\mathcal{E})}{|B|} \quad [3.71]$$

$$= \sum_{B \subseteq \Omega} m_y(B|\mathcal{E}) \left[ \frac{1}{|B|} \sum_{\omega \in B} C(\alpha, \omega) \right] \quad [3.72]$$

La stratégie de décision « bayésienne » consiste logiquement à choisir l'action de risque minimal. Ce principe général conduit à différentes règles de décision utiles en diagnostic, pour différents choix de l'ensemble des actions  $\mathcal{A}$  et de la fonction de coût  $\lambda$ . Le cas le plus simple est celui où les seules actions considérées sont les affectations aux classes :  $\mathcal{A} = \{\alpha_k\}_{k=1}^M$ , et où l'on pose

$$C(\alpha_j, \omega_k) = \begin{cases} 0 & \text{si } j = k \\ 1 & \text{sinon} \end{cases} \quad [3.73]$$

On a alors :

$$R(\alpha_k) = \sum_{j \neq k} p_y(\omega_j) = 1 - p_y(\omega_k) \quad \forall k \quad [3.74]$$

La règle de minimisation du risque pignistique revient dans ce cas à choisir la classe de probabilité pignistique maximale :

$$D(\mathbf{x}) = \alpha_k \Leftrightarrow p_y(\omega_k) \geq p_y(\omega_\ell) \quad \forall \ell \neq k \quad [3.75]$$

Ceci constitue un analogue crédibiliste de la règle de Bayes (cf. chapitre 1) et de la règle du maximum d'appartenance [3.38]. Incluons dans  $\mathcal{A}$  une action de rejet  $\alpha_r$  de coût constant  $c_r$

$$C(\alpha_r, \omega_k) = c_r \quad \forall k \quad [3.76]$$

On obtient la règle de décision suivante :

$$D(\mathbf{x}) = \begin{cases} \alpha_k & \text{si } p_y(\omega_k) \geq p_y(\omega_\ell) \quad \forall \ell \neq k \text{ et } p_y(\omega_k) \geq 1 - c_r \\ \alpha_r & \text{si } \max_{k=1, M} p_y(\omega_k) < 1 - c_r \end{cases} \quad [3.77]$$

Ceci revient à choisir la classe de probabilité pignistique maximale, à condition que cette probabilité soit supérieure à un seuil  $1 - c_r$ . Notons que, compte tenu du mode de calcul de la structure de croyance  $m_y(\cdot|\mathcal{E})$  défini par les équations [3.59]-[3.61], la masse  $m_y(\cdot|\mathcal{E})$  tend vers 1, et par conséquent les probabilités pignistiques  $p_y(\omega_k)$  tendent vers  $1/M$ , lorsque la distance de  $\mathbf{x}$  à l'ensemble d'apprentissage tend vers l'infini. Il en résulte que l'action de rejet  $\alpha_r$  dans la règle [3.77] se comporte à la fois comme un rejet d'ambiguïté et comme un rejet de distance, puisqu'elle tend à s'appliquer à la fois aux vecteurs formes situés dans une zone frontière entre plusieurs classes, et aux vecteurs « atypiques » situés loin de tout exemple d'apprentissage [DEN 97a]. Le taux de rejet peut être contrôlé par le seul paramètre  $c_r$ .

L'un des problèmes fondamentaux en diagnostic par reconnaissance de forme concerne la non exhaustivité de l'ensemble d'apprentissage, c'est-à-dire l'existence dans certaines applications de modes de fonctionnement inconnus ou pour lesquels on ne dispose d'aucune observation [DUB 90], [DUB 93], [DEN 98b]. Ce problème peut être traité dans le cadre défini précédemment en introduisant

**Tableau 3.2.** *Fonction de coût d'une règle avec rejet, dans le cas d'une connaissance incomplète des classes.*

	$\alpha_r$	$\alpha_1$	$\alpha_2$	$\dots$	$\alpha_{M-1}$	$\alpha_M$	$\alpha_0$
$\omega_1$	$c_r$	0	1	$\dots$	1	1	$c_0$
$\omega_2$	$c_r$	1	0	$\dots$	1	1	$c_0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\omega_{M-1}$	$c_r$	1	1	$\dots$	0	1	$c_0$
$\omega_M$	$c_r$	1	1	$\dots$	1	0	$c_0$
$\omega_0$	$c_r$	1	1	$\dots$	1	1	0

explicitement dans l'ensemble  $\Omega$  un élément supplémentaire  $\omega_0$  représentant globalement l'ensemble des classes inconnues :

$$\Omega = \{\omega_0, \omega_1, \dots, \omega_M\} \quad [3.78]$$

Aux actions définies précédemment vient par conséquent s'ajouter une action  $\alpha_0$  d'affectation à la classe  $\omega_0$ . En plus des coûts déjà définis, supposons que le choix de l'action  $\alpha_0$ , lorsque le système est dans un mode connu ( $y = \omega_k$ ,  $1 \leq k \leq c$ ) engendre un coût  $c_0$ , la fonction de coût étant alors spécifiée comme indiqué dans le tableau 3.2. On obtient alors les risques suivants :

$$R(\alpha_k) = 1 - p_y(\omega_k) \quad k \in \{1, \dots, M\} \quad [3.79]$$

$$R(\alpha_r) = c_r \quad [3.80]$$

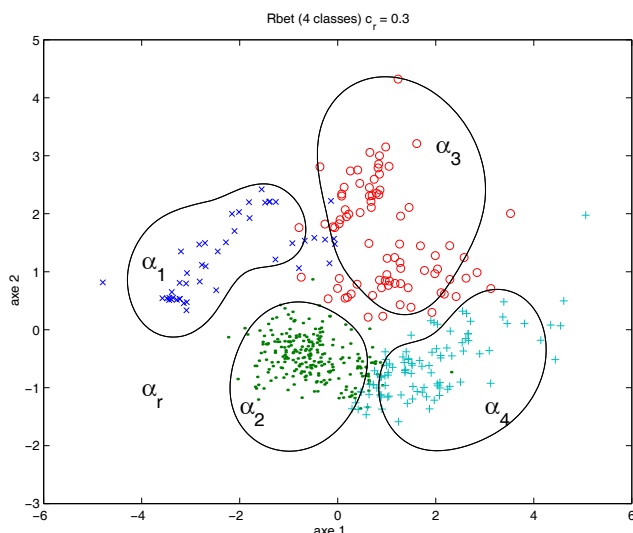
$$R(\alpha_0) = c_0[1 - p_y(\omega_0)] \quad [3.81]$$

Or, la probabilité pignistique affectée à  $\alpha_0$  n'étant due qu'à la masse  $m_y(\Omega|\mathcal{E})$ , on a

$$R(\alpha_0) = c_0 \left[ 1 - \frac{m_y(\Omega|\mathcal{E})}{M+1} \right] \quad [3.82]$$

Le risque pignistique associé au choix de l'action  $\alpha_0$  est donc une fonction décroissante de la masse  $m_y(\Omega|\mathcal{E})$ , et tend vers 1 lorsque la distance à l'ensemble d'apprentissage tend vers l'infini : la propension à choisir l'affectation à la classe inconnue est donc d'autant plus grande que l'entité à classer diffère davantage des exemples d'apprentissage.

EXEMPLE : Le comportement des règles de décision précédentes appliquées aux données de qualité d'eau (cf. paragraphe 3.2.4) est illustré sur les figures 3.4 et 3.5. Chaque classe a été résumée par deux prototypes optimisés par un algorithme connexionniste décrit dans les références [DEN 95a], [DEN 00c]. En l'absence de classe inconnue (figure 3.4), l'option de rejet est choisie entre les classes et dans les régions de faible densité d'observations, et correspond donc tantôt à un rejet d'ambiguïté, tantôt à un rejet de distance. L'ajout d'une classe



**Figure 3.4.** Données de qualité d'eau : régions de décision pour la règle du maximum de probabilité pignistique avec rejet.

inconnue au référentiel  $\Omega$  a pour effet de découpler ces deux types de rejet, la décision d'affectation à la classe inconnue étant prise lorsqu'aucun mode de fonctionnement connu ne semble crédible.

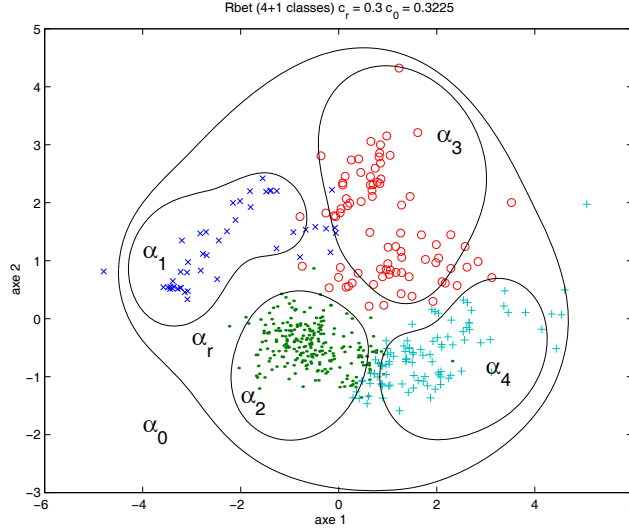
### 3.3.5. Extensions

Le formalisme des fonctions de croyance peut être étendu afin de représenter des croyances en des propositions floues (par exemple, «  $y$  est grand »), et des degrés de croyance flous («  $m(A) = \text{environ } 0.3$  »). Nous présentons brièvement ces deux extensions.

#### *Structures de croyance à éléments focaux flous*

L'idée de généraliser aux ensembles flous les notions relatives aux fonctions de croyance est due à Zadeh [ZAD 79], et a été explorée par de nombreux auteurs [YAG 82], [YAG 86], [SME 81], [YEN 90]. On définit une structure de croyance floue (SCF) comme une application de l'ensemble  $[0, 1]^\Omega$  des sous-ensembles flous de  $\Omega$  dans  $[0, 1]$ , telle que l'ensemble

$$\mathcal{F}(m) = \{F \in [0, 1]^\Omega \mid m(F) > 0\} \quad [3.83]$$



**Figure 3.5.** Données de qualité d'eau : régions de décision pour la règle du maximum de probabilité pignistique avec rejet et classe inconnue.

est fini, et

$$\sum_{F \in \mathcal{F}(m)} m(F) = 1 \quad [3.84]$$

La SCF  $m$  est dite normalisée si ses éléments focaux sont des sous-ensembles flous normalisés (de hauteur égale à 1).

La notion de degré de croyance  $\text{bel}_m(A)$  en une proposition floue «  $y$  est  $A$  », pour  $A \in [0, 1]^\Omega$  peut être définie par :

$$\text{bel}_m(A) = \sum_{F \in \mathcal{F}(m)} m(F) N(A|F) \quad [3.85]$$

$N(A|F)$  désigne la nécessité conditionnelle de  $A$  sachant  $F$  [ZAD 78] définie, dans le cas général où les ensembles  $A$  et  $F$  ne sont pas nécessairement normalisés [DUB 87], par :

$$N(A|F) = h_F - \max_{\omega \in \Omega} \min[\mu_A(\omega), 1 - \mu_F(\omega)] \quad [3.86]$$

$h_F$  désignant la hauteur de l'ensemble flou  $F$ . Remarquons que la quantité  $N(A|F)$  peut être interprétée comme un degré d'inclusion de  $F$  dans  $A$ , ce qui met peut-être mieux en évidence le lien entre les équations [3.48] et [3.85].

La somme conjonctive peut facilement être généralisée en remplaçant dans l'équation [3.49] l'intersection classique par un opérateur d'intersection entre ensembles flous.

Enfin, la probabilité pignistique associée à une SCF  $m$  peut également être définie [DEN 00b], [PET 99b], [PET 99c] par une formule analogue à l'équation [3.51], en utilisant la définition de la cardinalité pour les ensembles flous :

$$|B| = \sum_{\omega \in B} \mu_B(\omega) \quad [3.87]$$

et en étendant aux SCF l'opération de normalisation [YAG 96].

Ces différentes notions ont été appliquées par Zouhal et Dencœux [ZOU 96], [ZOU 97b],[DEN 01] au problème de l'apprentissage d'une règle de discrimination à partir d'exemples dont l'appartenance aux classes est partiellement définie sous la forme d'une distribution de possibilité, équivalente à une structure de croyance consonante [DUB 88]. L'application des équations [3.59]-[3.61] conduit alors à une SC à éléments focaux flous. L'intérêt de cette méthode a été démontré au travers d'une application à la détection de complexes K dans des électroencéphalogrammes [DEN 01].

#### *Masses de croyance floues*

Une étape supplémentaire dans la « fuzzification » du MCT peut être franchie en autorisant l'affectation aux différents éléments focaux (éventuellement flous) de *masses de croyance imprécises*, définies comme des intervalles ou des nombres flous [DEN 99b], [DEN 00a], [DEN 00b]. Cette généralisation est plus délicate que la précédente, car les différentes masses de croyance  $m(F)$ ,  $F \in \mathcal{F}(m)$  sont des variables interactives liées par la relation [3.84] ; on ne peut donc appliquer aux équations [3.49], [3.51] et [3.85] les principes simples de l'arithmétique d'intervalles ou de l'arithmétique floue. Dans [DEN 99b], nous avons abordé ce problème en définissant une structure de croyance à valeurs par intervalles comme un ensemble de SC dont les masses sont comprises entre certaines bornes ; nous avons montré que les notions de croyance et de plausibilité, de probabilité pignistique, d'opérateurs de combinaisons pouvaient être étendues rigoureusement dans ce contexte comme la solution de programmes linéaires ou quadratiques. Une généralisation supplémentaire a été effectuée dans [DEN 00a], [DEN 00b], où nous définissons une structure de croyance à valeurs floues comme un ensemble flou de SC, dont les masses sont contraintes par des distributions de possibilité, le principe d'extension de Zadeh [ZAD 75], [KLI 95], [BOU 95] permettant de retrouver l'ensemble des notions du MCT (degrés de croyance, probabilité pignistique, etc.).

Ce formalisme est immédiatement applicable au cas où, par suite d'incertitudes de mesures par exemple, certaines composantes  $x_j$  du vecteur forme  $\mathbf{x}$  ne sont que partiellement connues. Le cas où les variables  $x_j$  sont restreintes par des intervalles a été étudié dans [DEN 99b], puis étendu aux nombres flous dans

[DEN 00a], [DEN 00b]. L'approche retenue pour résoudre ce problème s'appuie sur les notions de SC imprécise (par intervalle ou floue), et consiste schématiquement à appliquer aux équations [3.59]-[3.61] le principe d'extension.

### 3.4. Conclusion

Nous avons dans ce chapitre décrit différentes approches pour la construction de systèmes de décision à partir de données. Ces approches reposent sur des modèles non probabilistes de l'incertain qui ne doivent pas être vus comme concurrents, mais comme complémentaires des modèles probabilistes. Les techniques de reconnaissance de formes floues, dont le développement a débuté il y a une trentaine d'années, se sont révélées particulièrement adaptées à la résolution de problèmes de diagnostic dans lesquels les modes de fonctionnement du système considéré ne peuvent être, sauf au prix d'une simplification excessive, définis de manière précise et non ambiguë. La théorie des fonctions de croyance, dont les applications en reconnaissance des formes ont commencé à être développées plus récemment, semble être à même de concilier certains avantages des probabilités et des ensembles flous, en permettant une représentation simultanée de l'imprécision et de l'incertitude dans la connaissance du mode de fonctionnement d'un système. La coexistence, dans la boîte à outils de l'ingénieur, de ces techniques avec les outils classiques de reconnaissance de formes statistique, ne peut que contribuer à étendre la démarche de diagnostic à des systèmes de plus en plus complexes.



## Bibliographie

- [APP 98] A. APPRIOU. Uncertain data aggregation in classification and tracking processes. In B. BOUCHON-MEUNIER, ed, *Aggregation and Fusion of imperfect information*, pages 231–260. Physica-Verlag, Heidelberg, 1998.
- [BEL 66] R. BELLMAN, L. KALABA ET L. A. ZADEH. Abstraction and pattern classification. *J. Math. Anal. Appl.*, 13 :1–7, 1966.
- [BÉR 86] M. BÉREAU. *Contribution de la théorie des sous-ensembles flous à la règle de discrimination des K plus proches voisins en mode partiellement supervisé*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1986.
- [BEZ 74] J.C. BEZDEK. Numerical taxonomy with fuzzy sets. *Journal of mathematical biology*, 1 :57–71, 1974.
- [BEZ 80] J.C. BEZDEK. A convergence theorem for the fuzzy isodata clustering algorithm. *IEEE Transactions on Pattern and Machine Intelligence*, 2(1) :1–8, 1980.
- [BEZ 81] J.C. BEZDEK. *Pattern Recognition with fuzzy objective function algorithm*. Plenum Press, New-York, 1981.
- [BEZ 86] J. C. BEZDEK, S. K. CHUAH ET D. LEEP. Generalized  $k$ -nearest neighbor rules. *Fuzzy Sets and Systems*, 18 :237–256, 1986.
- [BEZ 87] J.C. BEZDEK, R.J. HATHAWAY, M.J. SABIN ET W. T. TUCKER. Convergence theory for fuzzy c-means : Counterexamples and repairs. *IEEE Trans. Systems, Man and Cybernetics*, 17(5) :873–877, 1987.
- [BEZ 92] J. C. BEZDEK ET S. K. PAL. *Fuzzy models for pattern recognition*. IEEE Press, Piscataway, NJ, 1992.
- [BEZ 98] J. C. BEZDEK. Pattern analysis. In E. H. RUSPINI, P. P. BONISSONE ET W. PEDRYCZ, eds, *Handbook of Fuzzy Computation*, pages F6.1 :1–F6.6 :20. Institute of Physics Publishing, Bristol, 1998.
- [BEZ 99] J. C. BEZDEK, J. KELLER, R. KRISHNAPURAM ET N. R. PAL. *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer Academic Publishers, Boston, 1999.
- [BOU 93] E. BOUTLEUX ET B. DUBUISSON. Detection and supervision of the state evolution of the french telephone network. In *12th Int. Conference on Systems Science*, volume 4, pages 696–701, Vancouver, Canada, October 1993.
- [BOU 95] B. BOUCHON-MEUNIER. *La logique floue et ses applications*. Addison-Wesley, Paris, 1995.
- [BOU 96a] N. BOUDAUD, M. H. MASSON ET B. DUBUISSON. One-line diagnosis of a technological system : a fuzzy pattern recognition approach. In *13th IFAC Congress*, San Francisco, USA, August 1996.

- [BOU 96b] E. BOUTLEUX. *Diagnostic et suivi d'évolution de l'état d'un système par des fonctions d'appartenance bilatérales. Application au modèle du réseau téléphonique français*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1996.
- [BOU 97] N. BOUDAPOUD. *Conception d'un système de diagnostic adaptatif en ligne pour la surveillance des systèmes évolutifs*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1997.
- [BOU 99] N. BOUDAPOUD ET M. H. MASSON. Un système de diagnostic pour le suivi et la prédiction de la qualité de l'eau. In *Actes des XXXIes journées de statistiques*, pages 331–334, Grenoble, 1999.
- [CEL 89] G. CELEUX, E. DIDAY, G. GOVAERT, Y. LECHEVALLIER ET H. RALAMBONDRAINY. *Classification automatique des données*. Dunod, Paris, 1989.
- [CEL 95] G. CELEUX ET G. GOVAERT. Gaussian parsimonious clustering models. *Pattern Recognition*, 28(5) :781–793, 1995.
- [CRO 99] G. CRON. *Diagnostic par reconnaissance des formes floue d'un système dynamique et réparti. Application à la gestion en temps réel du trafic téléphonique français*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1999.
- [DEM 67] A. P. DEMPSTER. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, AMS-38 :325–339, 1967.
- [DEN 94] T. DENÈUX. Application of evidence theory to  $k$ -NN pattern classification. In E. S. GELSEMA ET L. N. KANAL, eds, *Pattern Recognition in Practice IV*, pages 13–24. Elsevier, Amsterdam, 1994.
- [DEN 95a] T. DENÈUX. An evidence-theoretic neural network classifier. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 3, pages 712–717, Vancouver, October 1995.
- [DEN 95b] T. DENÈUX. A  $k$ -nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05) :804–813, 1995.
- [DEN 97a] T. DENÈUX. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7) :1095–1107, 1997.
- [DEN 97b] T. DENÈUX. Function approximation in the framework of evidence theory : A connectionist approach. In *Proceedings of the 1997 International Conference on Neural Networks (ICNN'97)*, volume 1, pages 199–203, Houston, June 1997. IEEE.
- [DEN 98a] T. DENÈUX. Application du modèle des croyances transférables en reconnaissance de formes. *Traitement du Signal*, 14(5) :443–451, 1998.
- [DEN 98b] T. DENÈUX, M. MASSON ET B. DUBUISSON. Advanced pattern recognition techniques for system monitoring and diagnosis : A survey. *Journal Européen des Systèmes Automatisés (RAIRO-APII-JESA)*, 31(9-10) :1509–1539, 1998.
- [DEN 99a] T. DENÈUX. Modélisation de l'imprécis et de l'incertain en apprentissage supervisé par la théorie des fonctions de croyance (conférence invitée). In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 13–20, Valenciennes, October 1999. Cépaduès.
- [DEN 99b] T. DENÈUX. Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning*, 20 :79–111, 1999.
- [DEN 00a] T. DENÈUX. Allowing imprecision in belief representation using fuzzy-valued belief structures. In B. BOUCHON-MEUNIER, R. R. YAGER ET L. A. ZADEH, eds, *Information, Uncertainty and Fusion*, pages 269–281. Kluwer Academic Publishers, Boston, 2000.

- [DEN 00b] T. DENÈUX. Modeling vague beliefs using fuzzy-valued belief structures. *Fuzzy Sets and Systems*, 116(2) :167–199, 2000.
- [DEN 00c] T. DENÈUX. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics A*, 30(2) :131–150, 2000.
- [DEN 00d] T. DENÈUX ET M. SKARSTEIN-BJANGER. Induction of decision trees for partially classified data. In *Proceedings of SMC'2000*, pages 2923–2928, Nashville, TN, October 2000. IEEE.
- [DEN 01] T. DENÈUX ET L. M. ZOUHAL. Handling possibilistic labels in pattern classification using evidential reasoning. *To appear in Fuzzy Sets and Systems*, 2001.
- [DEV 98] A. DEVILLEZ, P. BILLAUDEL ET G. VILLERMAIN LECOLIER. Use of the fuzzy pattern matching method in a diagnosis module based on the pattern recognition. In *Proc. 2nd IMACS Int. Multiconf. CESA'98*, volume 4, pages 902–907, Nabeul-Hammamet, Tunisia, April 1998.
- [DEV 99] A. DEVILLEZ. *Contribution à la classification floue de données comportant des classes de forme quelconque. Application au développement d'un module d'aide à la décision*. PhD thesis, Université de Reims Champagne-Ardenne, Reims, 1999.
- [DID 74] E. DIDAY, A. SCHWEDER ET Y OK. The dynamic clusters method in pattern recognition. In *Proc. Int. Federation Information Processing Congress*, pages 691–709, New-York, 1974. Elsevier-North-Holland.
- [DUB 87] D. DUBOIS ET H. PRADE. An alternative approach to the handling of subnormal possibility distributions. *Fuzzy Sets and Systems*, 24 :123–126, 1987.
- [DUB 88] D. DUBOIS ET H. PRADE. Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.*, 4 :244–264, 1988.
- [DUB 90] B. DUBUISSON. *Diagnostic et Reconnaissance des Formes*. Hermès, Paris, 1990.
- [DUB 93] B. DUBUISSON ET M. MASSON. A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition*, 26(1) :155–165, 1993.
- [DUD 73] R. O. DUDA ET P. E HART. *Pattern Classification and Scene Analysis*. John Wiley and Sons, New-York, 1973.
- [DUN 74] J. C. DUNN. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3 :32–57, 1974.
- [ELO 00] Z. ELOUEDI, K. MELLOULI ET P. SMETS. Decision trees using the belief function theory. In *Proceedings of IPMU'2000*, volume 1, pages 141–148, Madrid, Spain, July 2000.
- [FRÉ 92] C. FRÉLICOT. *Un système adaptatif de diagnostic prédictif par reconnaissance des formes floues*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1992.
- [FRÉ 93] C. FRÉLICOT ET B. DUBUISSON. A posteriori ambiguity reject solving in fuzzy pattern classification using a multi-step predictor of membership vectors. In *Uncertainty in intelligent systems*. B. Bouchon-Meunier, L.Valverde, R. Yager, Elsevier Science, 1993.
- [FRÉ 95] C. FRÉLICOT, M.H MASSON ET B. DUBUISSON. Reject options in fuzzy classification rules. In *Proc. EUFIT'95*, volume III, pages 1495–1464, Aachen, Germany, August 1995.

- [FUK 90] K. FUKUNAGA. *Introduction to Statistical Pattern Recognition*. Academic Press, Boston, 1990.
- [GIO 98] A. GIORDANA ET L. SAITTA. Supervised learning. In E. H. RUSPINI, P. P. BONISSONE ET W. PEDRYCZ, eds, *Handbook of Fuzzy Computation*, pages F3.2 :1–D3.2 :7. Institute of Physics Publishing, Bristol, 1998.
- [GOV 75] G. GOVAERT. *Classification automatique et distances adaptatives*. PhD thesis, Université Paris VI, Paris, 1975.
- [GUS 79] E. E. GUSTAFSON ET W. KESSEL. Fuzzy clustering with a fuzzy covariance matrix. In *Proc. IEEE Conf. on Decision and Control*, pages 761–766, Piscataway, NJ, 1979. IEEE.
- [JOZ 83] A. JOZWIK. A learning scheme for a fuzzy k-nn rule. *Pattern Recognition Letters*, 1 :287–289, 1983.
- [KAS 96] N. K. KASABOV. *Foundations of neural networks, fuzzy systems, and knowledge engineering*. MIT Press, Cambridge, Ma, 1996.
- [KEL 85] J. M. KELLER, M. R. GRAY ET J. A. GIVENS. A fuzzy k-NN neighbor algorithm. *IEEE Trans. Syst. Man Cybern.*, SMC-15(4) :580–585, 1985.
- [KLI 95] G. J. KLIR ET B. YUAN. *Fuzzy sets and fuzzy logic. Theory and applications*. Prentice-Hall, Upper Saddle River, NJ., 1995.
- [KRI 93] R. KRISHNAPURAM ET J.M. KELLER. A possibilistic approach to clustering. *IEEE Trans. on Fuzzy Systems*, 1 :98–111, May 1993.
- [MAS 92] M. MASSON. *Contribution à l'élaboration d'une méthode de décision avec rejet par réseaux de neurones. Application à la surveillance des systèmes*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1992.
- [MAS 96] M.H. MASSON, B. DUBUISSON ET C. FRÉLICOT. Conception d'un module de reconnaissance des formes floues pour le diagnostic. *Journal Européen des Systèmes Automatisés (RAIRO-APII-JESA)*, pages 319–341, 1996.
- [MIL 95] S. MILLEMANN ET R. LENGELLÉ. Fuzzy supervised membership estimation using a multilayer perceptron. In *Proc. EUFIT'95*, volume I, pages 538–542, Aachen, Germany, August 1995.
- [NAU 98] D. NAUCK ET R. KRUSE. Neuro-fuzzy systems. In E. H. RUSPINI, P. P. BONISSONE ET W. PEDRYCZ, eds, *Handbook of Fuzzy Computation*, pages D2.1 :1–D2.10 :1. Institute of Physics Publishing, Bristol, 1998.
- [PAL 77] S. K. PAL ET D. MAJUMDER. Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 7 :625–629, 1977.
- [PEL 93a] M.-A. PELTIER. *Un système adaptatif de diagnostic d'évolution basé sur la reconnaissance des formes floues. Application au diagnostic du comportement d'un conducteur automobile*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1993.
- [PEL 93b] M.-A. PELTIER ET B. DUBUISSON. A human state detection system based on a fuzzy approach. In *Tooldiag'93, Int. Conf. on Fault Diagnosis*, pages 645–652, Toulouse, April 1993.
- [PEL 93c] M.A. PELTIER ET B. DUBUISSON. A fuzzy clustering algorithm based on the k-nearest neighbors rule for the detection of evolution. In *IEEE Int. Conference on Systems, Man and Cybernetics*, volume 4, pages 696–701, Le Touquet, France, October 1993.

- [PET 99a] S. PETIT-RENAUD ET T. DENGÈUX. Application de la théorie des fonctions de croyance en régression. In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 169–176, Valenciennes, October 1999. Cepaduès.
- [PET 99b] S. PETIT-RENAUD ET T. DENGÈUX. Handling different forms of uncertainty in regression analysis : a fuzzy belief structure approach. In A. HUNTER ET S. PEARSONS, eds, *Symbolic and quantitative approaches to reasoning and uncertainty (ECSQARU'99)*, pages 340–351, London, June 1999. Springer Verlag.
- [PET 99c] S. PETIT-RENAUD ET T. DENGÈUX. Regression analysis using fuzzy evidence theory. In *Proceedings of FUZZ-IEEE'99*, volume 3, pages 1229–1234, Seoul, August 1999.
- [RIC 98] C. RICHARD. *Une méthodologie pour la détection à structure imposée. Applications au plan temps-fréquence*. PhD thesis, Université de Technologie de Compiègne, 1998.
- [RIC 99] C. RICHARD ET R. LENGELLÉ. Data driven design and complexity control of time-frequency detectors. *Signal Processing*, 77 :37–48, 1999.
- [RUS 98a] E. H. RUSPINI, P. P. BONISSONE ET W. PEDRYCZ. *Handbook of Fuzzy Computation*. Institute of Physics Publishing, Bristol, 1998.
- [RUS 98b] E. H. RUSPINI ET F. ESTEVA. Interpretations of fuzzy sets. In E. H. RUSPINI, P. P. BONISSONE ET W. PEDRYCZ, eds, *Handbook of Fuzzy Computation*, pages B2.3 :1–B2.3 :13. Institute of Physics Publishing, Bristol, 1998.
- [SHA 76] G. SHAFER. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
- [SME 81] P. SMETS. The degree of belief in a fuzzy event. *Information Sciences*, 25 :1–19, 1981.
- [SME 90] P. SMETS. The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :447–458, 1990.
- [SME 93] P. SMETS. Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9 :1–35, 1993.
- [SME 94] P. SMETS ET R. KENNES. The Transferable Belief Model. *Artificial Intelligence*, 66 :191–243, 1994.
- [SME 97] P. SMETS. The normative representation of quantified beliefs by belief functions. *Artificial Intelligence*, 92(1–2) :229–242, 1997.
- [SME 98a] P. SMETS. Numerical representation of uncertainty. In D. M. GABBAY ET P. SMETS, eds, *Handbook of Defeasible reasoning and uncertainty management systems*, volume 3, pages 265–309. Kluwer Academic Publishers, Dordrecht, 1998.
- [SME 98b] P. SMETS. The Transferable Belief Model for quantified belief representation. In D. M. GABBAY ET P. SMETS, eds, *Handbook of Defeasible reasoning and uncertainty management systems*, volume 1, pages 267–301. Kluwer Academic Publishers, Dordrecht, 1998.
- [THO 53] R. L. THORNDIKE. Who belongs in the family? *Psychometrika*, 18 :267–276, 1953.
- [TRA 95] T. TRAUTMANN, M. CLÉMENT, T. DENGÈUX ET T. WITTIG. Application of intelligent techniques to river quality monitoring. In *Proceedings of EUFIT'95*, volume 3, pages 1603–1609, Aachen, August 1995.

- [YAG 82] R. R. YAGER. Generalized probabilities of fuzzy events from fuzzy belief structures. *Information Sciences*, 28 :45–62, 1982.
- [YAG 86] R. R. YAGER. Arithmetic and other operations on Dempster-Shafer structures. *Int. J. Man-Machines Studies*, 25 :357–366, 1986.
- [YAG 87] R. R. YAGER, S. OVCHINNIKOV, R. M. TONG ET H. T. NGUYEN. *Fuzzy sets and applications : Selected papers by L. A. Zadeh*. John Wiley and Sons, New-York, 1987.
- [YAG 96] R. R. YAGER. On the normalization of fuzzy belief structure. *International Journal of Approximate Reasoning*, 14 :127–153, 1996.
- [YEN 90] J. YEN. Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3) :559–569, 1990.
- [ZAD 65] L. A. ZADEH. Fuzzy sets. *Inform. Control*, 8 :338–353, 1965.
- [ZAD 75] L. A. ZADEH. The concept of a linguistic variable and its application to approximate reasoning (Part 1). *Information Sciences*, 8 :199–249, 1975.
- [ZAD 78] L. A. ZADEH. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.
- [ZAD 79] L. A. ZADEH. Fuzzy sets and information granularity. In R. K. RAGADE M. M. GUPTA ET R. R. YAGER, eds, *Advances in Fuzzy Sets Theory and Applications*, pages 3–18. North-Holland, Amsterdam, 1979.
- [ZIE 94] S. ZIEBA, M. SIDAHMED ET B. DUBUISSON. Tool wear monitoring and diagnosis in milling using vibration signals. In *SafeProcess'94*, pages 696–701, Espoo, Finland, June 1994.
- [ZIE 95] S. ZIEBA. *Une méthode de suivi d'un système évolutif. Application au diagnostic de la qualité d'usinage*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1995.
- [ZOU 95] L. M. ZOUHAL ET T. DENGEUX. Une méthode de discrimination non paramétrique basée sur la théorie de Dempster et Shafer. In *Actes du Quinzième Colloque GRETSI*, pages 689–692, Juan les Pins, Septembre 1995.
- [ZOU 96] L. M. ZOUHAL ET T. DENGEUX. Reconnaissance de formes floues par la théorie de Dempster et Shafer. In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 3–8, Nancy, December 1996. Cépaduès.
- [ZOU 97a] L. M. ZOUHAL. *Contribution à l'application de la théorie des fonctions de croyance en reconnaissance des formes*. PhD thesis, Université de Technologie de Compiègne, Compiègne, 1997.
- [ZOU 97b] L. M. ZOUHAL ET T. DENGEUX. Generalizing the evidence-theoretic  $k$ -NN rule to fuzzy pattern recognition. In *Proceedings of the Second International Symposium on Fuzzy Logic and Applications ISFL'97*, pages 294–300, Zurich, February 1997. ICSC Academic Press.
- [ZOU 98] L. M. ZOUHAL ET T. DENGEUX. An evidence-theoretic  $k$ -NN rule with parameter optimization. *IEEE Trans. on Systems, Man and Cybernetics C*, 28(2) :263–271, 1998.

# Index

- $K$  plus proches voisins, 14, 26
- apprentissage
  - non supervisé, 8
  - supervisé, 15, 20, 23, 25, 31
- classification floue, 11, 13, 14
- conorme triangulaire, 8, 14
- ensemble d'apprentissage, 14, 15, 20, 23, 25
- FCM, 9, 11
- flou (ensemble), 5, 7
- fonction
  - d'appartenance, 13–15
  - de croyance, 21, 23
- norme triangulaire, 8
- partition floue, 8–10, 12, 13
- pignistique (probabilité), 22, 26, 31
- prototype, 9, 13, 14, 26
- rejet, 3, 16–18, 27, 29
- structure de croyance, 21, 26, 27, 30, 31