Reconnaissance de Formes Floues par la Théorie de Dempster et Shafer *

L. M. Zouhal[1,2] et T. Denœux[1]
[1] Université de Technologie de Compiègne - U.R.A CNRS 817 Heudiasyc
BP 529 - F-60205 Compiègne cedex - France
[2] Lyonnaise des Eaux
Centre International des Techniques Informatiques
email: Meriem.Zouhal/Thierry.Denoeux@hds.utc.fr

Résumé

Une approche de la discrimination en reconnaissance des formes consiste à considérer les éléments d'un ensemble d'apprentissage comme des sources d'information relatives au classement d'un nouveau vecteur. Les différentes sources sont représentées par des structures de croyances et combinées par la règle de Dempster. Dans cet article, nous proposons d'étendre cette approche au cas où l'appartenance d'un vecteur d'apprentissage aux classes est caractérisée par un ensemble flou. Deux méthodes sont proposées, reposant respectivement sur la théorie des possibilités, et sur une généralisation de la théorie des fonctions de croyance autorisant l'existence d'éléments focaux flous. Ces méthodes sont comparées expérimentalement sur deux exemples.

1 Introduction

En Reconnaissance de Formes, la prise de décision est souvent nécessaire malgré la nature incertaine des informations disponibles. La richesse et la complexité des données, sans cesse croissantes, nécessitent l'utilisation de méthodes de plus en plus performantes. Dans ce contexte, une nouvelle méthode des k plus proches voisins (k-ppv) basée sur la théorie de Dempster et Shafer (D-S) a récemment été proposée [1]. Dans cette méthode, la classification d'un nouveau vecteur est basée sur la combinaison des informations apportées par chacun de ses plus proches voisins dans l'ensemble d'apprentissage. Chaque voisin est considéré comme accréditant certaines hypothèses concernant la classe d'appartenance du vecteur à classer. A cette source d'information est associée une structure de croyance[1] définie en fonction de la distance entre les deux vecteurs. Les différentes structures sont ensuite combinées par la règle de Dempster [5].

Dans cet article, une extension de cette méthode aux ensembles flous est proposée. Chaque vecteur de l'ensemble d'apprentissage n'est plus supposé associé à une classe unique mais à toutes les classes existantes par le biais d'une fonction d'appartenance. Ceci permet de considérer chaque voisin d'un vecteur à classer comme étiqueté par un sous-ensemble flou de l'ensemble Ω des classes.

Deux approches ont été développées. La première repose sur le lien entre les notions de sous-ensemble flou, de distribution de possibilité et de fonction de croyance consonante [2]. A chaque sous-ensemble flou F représentant l'appartenance aux classes d'un vecteur d'apprentissage est associée une structure de croyance dont les éléments focaux sont les α-coupes de F. La combinaison des informations peut être effectuée soit par la règle de Dempster soit par un opérateur de combinaison possibiliste appliqué aux distributions de possibilité. La seconde approche est basée sur une généralisation de la théorie de D-S aux ensembles flous proposée par Yager [7, 8]. Elle consiste à associer à chaque vecteur d'apprentissage une structure de croyance dont les éléments focaux sont des sous-ensembles flous de Ω. Les différentes structures sont ensuite combinées par la règle de Dempster généralisée [8].

Rencontres Francophones sur la Logique Floue et ses Applications, pages 3–8, Nancy, Decembre 1996. Cépaduès.

1. Nous adopterons le terme “structure de croyance” pour désigner la fonction de masse appelée *basic probability assignment* par Shafer [5] et *basic belief assignment* par Smets [6].

La structure de l'article est la suivante. Nous commençons par rappeler le principe de l'application de la théorie de D-S en Reconnaissance de Formes. Les deux méthodes proposées sont ensuites décrites successivement puis comparées expérimentalement à d'autres règles de discrimination basées sur le principe des k plus proches voisins. Quelques remarques finales sont données en conclusion.

2 Théorie de D-S en Reconnaissance des Formes

Soient $\mathcal{X}$ un ensemble d'apprentissage composé de N vecteurs répartis en M classes $\Omega = \{\omega_1, \ldots, \omega_M\}$, et $\boldsymbol{x}^s$ un nouveau vecteur forme à classer. Soit ϕ_k^s l'ensemble des k-ppv de $\boldsymbol{x}^s$ dans $\mathcal{X}$. Chaque voisin $\boldsymbol{x}^i$ appartenant à ϕ_k^s est supposé caractérisé par une étiquette $L^i = q \in \{1, \ldots, M\}$ définissant de manière certaine et non ambigüe son appartenance à la classe ω_q [2]. Cette information peut être représentée par une structure de croyance $\mathrm{m}^{s,i} : 2^\Omega \mapsto [0,1]$, d'éléments focaux $\{\omega_q\}$ et Ω, définie par :

$$\begin{aligned} \mathrm{m}^{s,i}(\{\omega_q\}) &= \alpha_q^{s,i} \qquad (1)\\ \mathrm{m}^{s,i}(\Omega) &= 1 - \alpha_q^{s,i} \qquad (2) \end{aligned}$$

et $\mathrm{m}^{s,i}(A) = 0$ pour tout $A \in 2^\Omega \setminus \{\Omega, \{\omega_q\}\}$. Le terme $\alpha_q^{s,i}$ est une fonction décroissante de la distance $d^{s,i}$ entre les vecteurs $\boldsymbol{x}^s$ et $\boldsymbol{x}^i$:

$$\alpha_q^{s,i} = \alpha_0 \exp(-\gamma_q (d^{s,i})^2) \qquad (3)$$

γ_q étant un paramètre caractérisant la classe ω_q et α_0 un paramètre fixé. La règle de combinaison de Dempster [5] permet de construire une nouvelle distribution de masse m en combinant les distributions $\mathrm{m}^{s,1}, \ldots, \mathrm{m}^{s,k}$ associées aux k plus proches voisins de $\boldsymbol{x}^s$. Les éléments focaux de m sont les classes représentées dans ϕ_k^s, et Ω. La crédibilité et la plausibilité de ω_q peuvent être calculées en fonction de m :

$$\begin{aligned} \mathrm{bel}(\{\omega_q\}) &= \mathrm{m}(\{\omega_q\}) \qquad (4)\\ \mathrm{pl}(\{\omega_q\}) &= \mathrm{m}(\{\omega_q\}) + m(\Omega) \qquad (5) \end{aligned}$$

2. Le cas plus général dans lequel l'appartenance des vecteurs d'apprentissage aux classes est incertaine et caractérisée par une structure de croyance a également été abordé dans [1].

Toute distribution de probabilité P telle que $\mathrm{bel}(A) \leq P(A) \leq \mathrm{pl}(A)$ pour toute partie A de Ω est dite compatible avec m [5]. Un cas particulier est la distribution de probabilité "pignistique" [6] BetP obtenue en répartissant uniformément la masse $m(\Omega)$ entre toutes les classes :

$$\mathrm{BetP}(\{\omega_q\}) = \mathrm{m}(\{\omega_q\}) + \frac{\mathrm{m}(\Omega)}{M} \qquad (6)$$

Le vecteur forme $\boldsymbol{x}^s$ est finalement affecté à la classe de plus grande crédibilité, qui est également dans ce cas celle de plus grandes plausibilité et probabilité pignistique. Les paramètres γ_q intervenant dans l'équation 3 peuvent être déterminés par minimisation d'un critère d'erreur [10, 11].

3 Approche par la Théorie des Possibilités

3.1 Structures de croyance et mesures de possibilité

Une première voie permettant de relier les notions d'ensemble flou et de fonction de croyance passe par la théorie des Possibilités. En effet, une mesure de possibilité est formellement équivalente à une mesure de plausibilité dont les éléments focaux sont emboîtés ; la mesure de possibilité associée est alors une fonction de croyance dite consonante [2]. Or, une distribution de possibilité peut être définie à partir d'un ensemble flou. Un lien entre ensemble flou et structure de croyance consonante est donc établi. Formellement, la construction est la suivante.

Soit F un sous-ensemble flou normalisé du référentiel Ω supposé fini et μ_F sa fonction d'appartenance. La distribution de possibilité π induite par F [9] est définie par :

$$\pi(\omega) = \mu_F(\omega) \quad \forall \omega \in \Omega \qquad (7)$$

Les mesures de possibilité et de nécessité associées sont définies respectivement par

$$\begin{aligned} \Pi(A) &= \max_{\omega \in A} \pi(\omega) \qquad (8)\\ N(A) &= \min_{\omega \in A} (1 - \pi(\omega)) \qquad (9) \end{aligned}$$

pour tout $A \subseteq \Omega$. Les fonctions Π et N sont respectivement des mesures de plausibilité et de crédibilité dont les éléments focaux sont les α-coupes de F. De plus, soient $\pi_1 = 1 > \pi_2 > \ldots > \pi_r$

les valeurs distinctes prises par la distribution π, et $\pi_{r+1} = 0$ par convention. Alors, la structure de croyance associée à Π est définie par [2] :

$$\mathrm{m}(F_i) = \pi_i - \pi_{i+1} \quad i = 1, \ldots, r \tag{10}$$

F_i désignant la π_i-coupe de F. L'existence d'un lien formel entre les notions de mesure de possibilité et de structure de croyance permet d'envisager plusieurs modes de combinaison selon que l'on se place du point de vue de la théorie des possibilités ou de celui de la théorie des fonctions de croyance [3]. En effet, on peut définir à partir de deux distributions de possibilité π_1 et π_2 une nouvelle distribution π par une opération de type conjonctif en posant :

$$\pi(\omega) = \frac{\pi_1(\omega) * \pi_2(\omega)}{\max_{\omega \in \Omega} \ \pi_1(\omega) * \pi_2(\omega)} \tag{11}$$

où $*$ est une norme triangulaire quelconque, par exemple l'opérateur min. Mais il est également possible de combiner par la règle de Dempster les structures de croyance m_1 et m_2 associées à π_1 et π_2. Remarquons qu'en général la structure de croyance m ainsi obtenue n'est plus consonante et ne permet donc pas de définir une distribution de possibilité équivalente.

3.2 Application

Chaque voisin $\boldsymbol{x}^i$ d'un vecteur forme à classer $\boldsymbol{x}^s$ est maintenant supposé caractérisé par un vecteur $\boldsymbol{\mu}^i = (\mu_1^i, \ldots, \mu_M^i)^t$ où μ_j^i est le degré d'appartenance de l'individu $\boldsymbol{x}^i$ à la classe floue ω_j. On définit donc pour chaque vecteur d'apprentissage $\boldsymbol{x}^i$ un sous-ensemble flou F^i de Ω [9] en posant :

$$F^i = \frac{\mu_1^i}{\omega_1} + \ldots + \frac{\mu_M^i}{\omega_M} \tag{12}$$

L'ensemble flou F^i induit une structure de croyance consonante m^i. On définit à partir de m^i et de la distance $d^{s,i}$ entre $\boldsymbol{x}^s$ et $\boldsymbol{x}^i$ une nouvelle structure de croyance $\mathrm{m}^{s,i}$ caractérisant l'information apportée par le couple $(\boldsymbol{x}^i, F^i)$ relativement à la classe de $\boldsymbol{x}^s$:

$$\mathrm{m}^{s,i}(A) = \alpha^{s,i} m^i(A) \quad \forall A \subset \Omega \tag{13}$$

$$\mathrm{m}^{s,i}(\Omega) = 1 - \alpha^{s,i} \sum_{A \neq \Omega} m^i(A) \tag{14}$$

où $\alpha^{s,i}$ est un terme décroissant en fonction de la distance :

$$\alpha^{s,i} = \alpha_0 \exp(-\gamma (d^{s,i})^2) \tag{15}$$

Les éléments focaux de $\mathrm{m}^{s,i}$ sont les π_i-coupes du sous-ensemble flou F^i.

Deux procédures peuvent être envisagées pour combiner les informations apportées par les k-ppv de $\boldsymbol{x}^s$:

1. combinaison des structures de croyance $\mathrm{m}^{s,i}$ pour les k-ppv de $\boldsymbol{x}^s$ par la règle de Dempster (méthode 1);

2. combinaison des distributions de possibilité $\pi^{s,i}$ associées aux structures de croyance $\mathrm{m}^{s,i}$ (méthode 2); la distribution de possibilité π résultante induit alors une structure de croyance consonante.

Dans les deux cas, le paramètre γ apparaissant dans l'équation 15 peut être déterminé par minimisation d'un critère d'erreur moyen calculé sur l'ensemble d'apprentissage. Selon la procédure utilisée, nous proposons d'utiliser deux critères d'erreur distincts $E_a(\boldsymbol{x})$ et $E_b(\boldsymbol{x})$ définis respectivement par :

$$E_a(\boldsymbol{x}) = \frac{1}{2} \sum_{j=1}^{M} (\mathrm{BetP}(\{\omega_j\}) - \mu_j)^2 \tag{16}$$

$$E_b(\boldsymbol{x}) = \frac{1}{2} \sum_{j=1}^{M} (\pi(\omega_j) - \mu_j)^2 \tag{17}$$

μ_j étant le degré d'appartenance du vecteur d'apprentissage $\boldsymbol{x}$ à la classe ω_j et BetP la distribution de probabilité pignistique associée à la structure de croyance obtenue pour $\boldsymbol{x}$ à partir de ses k plus proches voisins. Remarquons que le critère E_a, qui généralise le critère proposé dans [10] pour le cas non flou, est seul compatible avec la méthode 1, qui conduit à une fonction de croyance non consonante.

4 Approche par éléments focaux flous

4.1 Généralisation de la théorie de D-S aux ensembles flous

Dans la théorie de D-S, les masses de croyance sont affectées à des sous-ensembles classiques du référentiel Ω. Plusieurs généralisations de cette théorie aux ensembles flous ont été proposées. Nous retiendrons ici l'approche proposée par Yager [7, 8]. Soit $\mathcal{F}$ l'ensemble des sous-ensembles flous de Ω. Une structure de croyance floue est une fonction m de $\mathcal{F}$ dans $[0, 1]$ vérifiant les propriétés suivantes :

1. l'ensemble $\mathcal{F}'$ des éléments A de $\mathcal{F}$ tels que $\mathrm{m}(A) > 0$ (appelés éléments focaux de m) est fini;
2. $\sum_{A \in \mathcal{F}'} \mathrm{m}(A) = 1$.
3. $\forall A \in \mathcal{F}, \max_{\omega \in \Omega} \mu_A(\omega) < 1 \Rightarrow \mathrm{m}(A) = 0$.

Une fonction m ne vérifiant pas la condition 3 est appelée "pseudo-structure de croyance" par Yager [8]. On étend ensuite les notions de crédibilité et de plausibilité d'un élément quelconque A de $\mathcal{F}$ en posant :

$$\begin{aligned} \mathrm{bel}(A) &= \sum_{B \in \mathcal{F}'} \mathrm{m}(B) \mathrm{Inc}(B, A) && (18) \\ \mathrm{pl}(A) &= \sum_{B \in \mathcal{F}'} \mathrm{m}(B) \mathrm{Int}(B, A) && (19) \end{aligned}$$

où $\mathrm{Inc}(B, A)$ et $\mathrm{Int}(B, A)$ sont respectivement des mesures d'inclusion de B dans A et d'intersection de B et A, définies par exemple par :

$$\begin{aligned} \mathrm{Inc}(B, A) &= \min_{\omega \in \Omega} [\mu_A(\omega) \vee \mu_{\bar{B}}(\omega)] && (20) \\ \mathrm{Int}(B, A) &= \max_{\omega \in \Omega} [\mu_A(\omega) \wedge \mu_B(\omega)] && (21) \end{aligned}$$

où $\bar{B}$ désigne le complémentaire de B. La règle de Dempster sans normalisation se généralise immédiatement de la manière suivante. Etant données deux structures de croyance floues m_1 et m_2 définies sur $\mathcal{F}$, ont pose :

$$\mathrm{m}(A) = \sum_{B, C \in \mathcal{F}', B \cap C = A} \mathrm{m}_1(B) \mathrm{m}_2(C) \quad \forall A \in \mathcal{F} \tag{22}$$

La fonction m est en général une pseudo-structure de croyance, que l'on peut transformer en structure de croyance m^* par l'opération de normalisation suivante :

– on remplace chaque élément focal F_i de m par un élément focal normalisé E_i défini par $E_i(x) = F_i(x)/h_i$, h_i étant la hauteur de F_i;

– on note $T = \sum_i \mathrm{m}(F_i)(1 - h_i)$;

– on pose $\mathrm{m}^*(E_i) = \mathrm{m}(F_i) h_i / (1 - T)$.

Remarquons que l'opération de normalisation fait perdre la propriété d'associativité de la règle de combinaison. Lorsque l'on souhaite combiner plusieurs structures de croyance $\mathrm{m}^1, \ldots, \mathrm{m}^k$, il faut donc en général calculer dans un premier temps la somme non normalisée de ces k structures, puis effectuer une seule fois l'opération de normalisation.

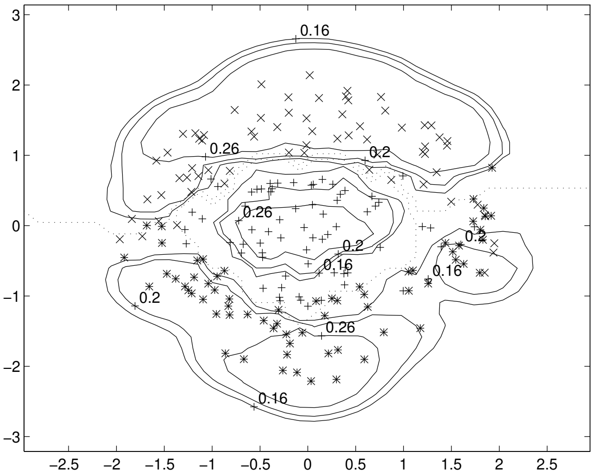

FIG. 1 - *Données non gaussiennes et courbes de niveau de la fonction de crédibilité obtenue par la méthode 3 (éléments focaux flous).*

4.2 Application

On suppose comme précédemment chaque vecteur d'apprentissage $\boldsymbol{x}^i$ étiqueté par un sous-ensemble flou F_i de Ω, et on considère un nouveau vecteur $\boldsymbol{x}^s$ à classer. A chaque sous-ensemble flou $F^i, i = 1, \ldots, k$ est maintenant associée une structure de croyance floue $\mathrm{m}^{s,i} : \mathcal{F} \mapsto [0, 1]$ définie par :

$$\begin{aligned} \mathrm{m}^{s,i}(F^i) &= \alpha^{s,i} && (23) \\ \mathrm{m}^{s,i}(\Omega) &= 1 - \alpha^{s,i} && (24) \end{aligned}$$

et $\mathrm{m}^{s,i}(A) = 0$ pour tout $A \in \mathcal{F} \backslash \{\Omega, F^i\}$. Le terme $\alpha^{s,i}$ est supposée comme précédemment fonction décroissante de la distance $d^{s,i}$ entre les vecteurs $\boldsymbol{x}^s$ et $\boldsymbol{x}^i$.

La masse de croyance m associée à $\boldsymbol{x}^s$ est déterminée par combinaison des différentes masses $\mathrm{m}^{s,i}$ selon la règle de Dempster généralisée. On calcule ensuite la crédibilité et la plausibilité de chaque singleton de Ω à l'aide des équations 18 et 19. Le vecteur $\boldsymbol{x}$ est finalement affecté à la classe de plus grande crédibilité, ou à celle de plus grande plausibilité.

Le paramètre γ intervenant dans le calcul de $\alpha^{s,i}$ peut être déterminé par apprentissage selon une procédure généralisant celle décrite dans

[10, 11] dans le cas non flou. Pour chaque vecteur d'apprentissage $\boldsymbol{x}$ de vecteur d'appartenance $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)^t$, on calcule comme indiqué précédemment une fonction de plausibilité pl et on définit un critère d'erreur $E(\boldsymbol{x})$:

$$E(\boldsymbol{x}) = \frac{1}{2}\sum_{j=1}^{M}(\mathrm{pl}(\{\omega_j\}) - \mu_j)^2 \qquad (25)$$

La valeur de γ est déterminée par minimisation de l'erreur moyenne sur l'ensemble d'apprentissage [10].

5 Exemples

Nous présentons à titre d'exemple les résultats obtenus par les trois méthodes décrites dans cet article pour deux jeux de données. Le premier est représenté sur la figure 1. Il s'agit de données non gaussiennes composées de 3 classes en dimension 2. Le deuxième jeu de données se compose de 3 classes gaussiennes de paramètres :

$$\begin{array}{ll} \boldsymbol{\xi}_1 = (1,1,1)^t & \Sigma_1 = I \\ \boldsymbol{\xi}_2 = (-1,1,0)^t & \Sigma_2 = 4I \\ \boldsymbol{\xi}_3 = (0,-1,1)^t & \Sigma_3 = 3I \end{array}$$

Dans les deux cas, l'ensemble d'apprentissage est de taille $N = 90$ et l'ensemble de test se compose de 600 exemples. Pour chaque vecteur d'apprentissage $\boldsymbol{x}$, on définit un degré d'appartenance μ_i à la classe ω_i égal à :

$$\mu_i = \frac{1}{1 + d(\boldsymbol{x}, \bar{\boldsymbol{x}}_k^i)} \qquad (26)$$

où $\bar{\boldsymbol{x}}_k^i$ représente le centre de gravité des k-ppv de $\boldsymbol{x}$ appartenant à la classe ω_i. Ce prétraitement de l'ensemble d'apprentissage revient donc à relativiser l'appartenance aux classes de chaque vecteur en tenant compte de l'information locale.

Les taux d'erreur de test obtenus pour chacune des trois méthodes proposées ont été comparés pour différentes valeurs de k à ceux obtenus par la règle des k plus proches voisins avec vote majoritaire, la règle des k-ppv floue introduite par Keller [4] et par la règle basée sur la théorie de Dempster et Shafer sans prise en compte du flou (Figure 2). Pour ces deux exemples, les trois méthodes proposées s'avèrent équivalentes en termes de performances, et nettement supérieures aux deux approches “classiques”. Ceci montre l'intérêt d'un étiquetage flou de l'ensemble d'apprentissage et des méthodes présentées pour prendre en compte un tel étiquetage.

6 Conclusion

Différentes méthodes permettant de prendre en compte un étiquetage flou de l'ensemble d'apprentissage dans le cadre de la théorie des fonctions de croyance ont été proposées. Ces méthodes reposent sur le lien entre les notions de structures de croyance consonante et de distribution de possibilité d'une part, et sur une généralisation des structures de croyance à des éléments focaux flous d'autre part. Bien que les performances obtenues par chacune de ces méthodes semblent expérimentalement équivalentes pour les deux exemples étudiés, la méthode reposant sur des éléments focaux flous semble plus riche car elle permet potentiellement de prendre en compte à la fois l'imprécision et l'incertitude sur la classe d'un vecteur d'apprentissage en lui associant une structure de croyance floue.

Références

[1] T. Denœux, A k-nearest neighbor classification rule based on Dempster-Shafer Theory, IEEE Transactions on Systems, Man and Cybernetics 25(5) (1995) 804-813.

[2] D. Dubois and H. Prade, On several representations of an uncertainty body of evidence, M. M. Gupta and E. Sanchez (Editors), Fuzzy Information and Decision Processes, North-holland, New-York (1982) 167-181.

[3] D. Dubois and H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, International Journal of General Systems 12 (1986) 193-226.

[4] J. M. Keller and M. R. Gray and J. A. Givens, A fuzzy k-NN neighbor algorithm, IEEE Transactions on Systems, Man and Cybernetics 15(4) (1985) 580-585.

[5] G. Shafer, A mathematical theory of evidence, Princeton N.J, Princeton University Press (1976).

[6] P. Smets and R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191-234.

[7] R. R. Yager, Generalized probabilities of fuzzy events from fuzzy belief structures, Information Sciences 28 (1982) 45-62.

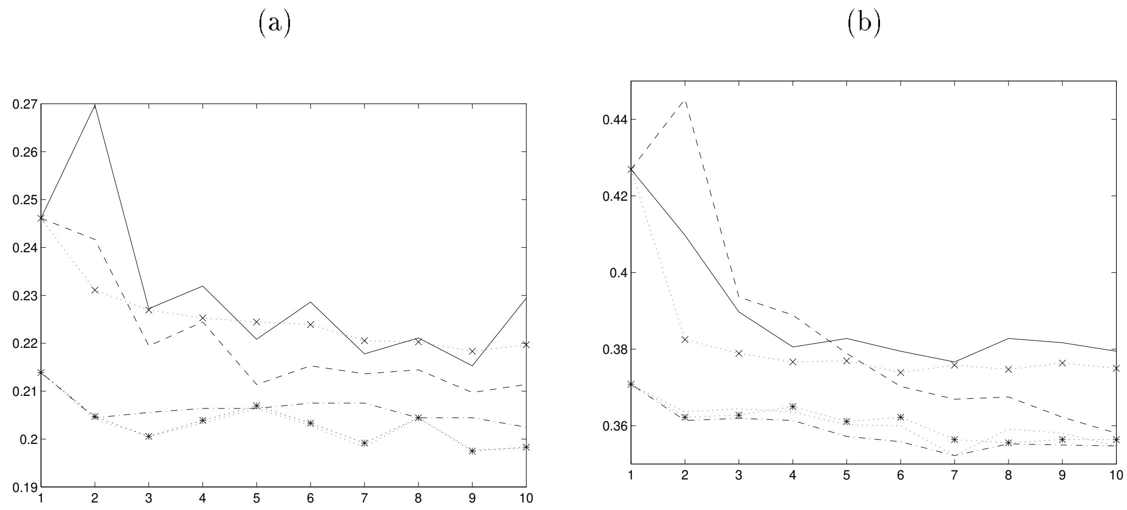

FIG. 2 - *Taux d'erreur de test pour les règles: k-ppv classique (-), k-ppv floue (.×), k-ppv basée sur la théorie de D-S (- -), k-ppv méthode 1 (..), k-ppv méthode 2 (.∗) et k-ppv méthode (-.), pour différentes valeurs de k; (a): données non gaussiennes; (b): données gaussiennes.*

[8] R. R. Yager and D. P. Filev, Including probabilistic uncertainty in fuzzy logic controller modeling using Dempster-Shafer theory, IEEE Transactions on Systems, Man and Cybernetics 25 (1995) 1221-1230.

[9] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy sets and systems 1 (1978) 3-28.

[10] L. M. Zouhal and T. Denœux, An Adaptive k-NN rule based on Dempster-Shafer theory, Proc. of the 6th Int. Conf. on Computer Analysis of Images and Patterns (CAIP'95), Springer Verlag, Prague (1995), 310-317.

[11] L. M. Zouhal et T. Denœux, Une méthode de discrimination non paramétrique basée sur la théorie de Dempster et Shafer, Actes du Quinzième Colloque GRETSI, Juan les Pins (1995) 689-692.

[12] L. M. Zouhal and T. Denœux, A comparison between fuzzy and evidence-theoretic k-NN rules for pattern recognition, Proceedings of EUFIT'95, Aachen 3 (1995) 1319-1325.