Generalizing the evidence-theoretic k-NN rule to Fuzzy Pattern Recognition *

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

Abstract

A pattern classifier using neighborhood information in the framework of Dempster-Shafer theory was recently proposed. This method consists in considering each neighbor of a pattern $\boldsymbol{x}$ to be classified as an item of evidence supporting certain hypotheses concerning the class membership of $\boldsymbol{x}$. This information is represented by a belief structure defined as a function of the distance between the two vectors. The evidence of the k nearest neighbors is then pooled by means of Dempster's rule of combination. In this paper, this approach is extended to handle the more general situation in which training patterns are assumed to have some degree of membership to each class. Two methods are described. The first one exploits the relationship between fuzzy sets, possibility distributions and consonant belief structures. The second approach is based on a generalization of the Dempster-Shafer theory to fuzzy sets. These methods are compared experimentally with the voting, evidence-theoretic and fuzzy k nearest neighbors rules.

1 Introduction

In Pattern Recognition, a lot of effort has been devoted to the development of efficient classification procedures. In real-world applications of statistical pattern recognition, no prior knowledge about class-conditional probability distributions is generally available, and classification of input patterns has to be based on recorded data only. The imperfectness of the available information must be recognized by taken into account various types of uncertainty and imprecision in the classification procedure. The Dempster-Shafer and Fuzzy sets theories can both be used to resolve uncertainty of information in a decision process. In [1], the classification problem was addressed from the point of view of the Dempster-Shafer theory of evidence [5]. In this approach, each of the nearest neighbors of a pattern to be classified is considered as an item of evidence supporting certain hypotheses concerning the class membership of that pattern. This information is provided in the form of a belief structure, defined as a function of the distance between the pattern under consideration and its neighbor.

In this paper, this approach is extended to handle the more general situation in which training patterns are assumed to have some degree of membership to each class. Class labels are then defined as fuzzy subsets of the reference set Ω. Two methods are described. The first one exploits the relationship between fuzzy sets, possibility distributions and consonant belief structures [2]. A fuzzy subset F representing the fuzzy class membership of a training pattern is converted into a consonant belief structure whose focal elements are the α-cuts of F. The evidence of the k nearest neighbors (k-NN) of a pattern

*In *Proceedings of the Second International Symposium on Fuzzy Logic and Applications ISFL'97*, pages 294–300, Zurich, February 1997. ICSC Academic Press.

to be classified is then pooled using either Dempster's rule of combination [5] or a possibilistic rule such as fuzzy set intersection [3]. The second approach is based on a generalization of the Dempster-Shafer theory to fuzzy sets [7, 8]. It consists in representing the evidence of each training pattern by a belief structure with fuzzy focal elements. The combination of the different structures is then performed using the generalized Dempster's rule.

In the following section, a brief overview of the evidence-theoretic k-NN rule is provided. The two proposed generalizations are then described. Finally, several sets are used to perform an experimental comparison of these methods with the voting, evidence-theoretic and fuzzy k-NN rules.

2 The evidence-theoretic k-NN rule

Let us consider a pattern $\boldsymbol{x}^s$ to be classified, and ϕ_k^s the set of its k nearest neighbors in a training set $\mathcal{X}$. Each $\boldsymbol{x}^i \in \phi_k^s$, is assumed to possess a class label $L^i = q \in \{1, \ldots, M\}$ indicating with certainty its membership to ω_q. For any $\boldsymbol{x}^i \in \phi_k^s$, the knowledge that $L^i = q$ can be interpreted as a piece of evidence that increases our belief that $\boldsymbol{x}^s$ also belongs to ω_q [1]. This item of evidence can be represented by a *belief structure* (BS) $\mathrm{m}^{s,i}$ over the set Ω of classes, i.e. a function $\mathrm{m}^{s,i} : 2^\Omega \longmapsto [0,1]$ verifying $\mathrm{m}^{s,i}(\emptyset) = 0$ and $\sum_{A\subseteq\Omega} \mathrm{m}^{s,i}(A) = 1$. The function $\mathrm{m}^{s,i}$ distributes a fraction $\alpha_q^{s,i}$ of the unit mass to the singleton $\{\omega_q\}$, the rest being committed to Ω:

$$\mathrm{m}^{s,i}(\{\omega_q\}) = \alpha_q^{s,i} \tag{1}$$

$$\mathrm{m}^{s,i}(\Omega) = 1 - \alpha_q^{s,i} \tag{2}$$

and $\mathrm{m}^{s,i}(A) = 0$ for all $A \in 2^\Omega \setminus \{\Omega, \{\omega_q\}\}$. The quantity $\alpha_q^{s,i}$ is defined as a decreasing function of the distance $d^{s,i}$ between $\boldsymbol{x}^s$ and $\boldsymbol{x}^i$:

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

where γ_q is a parameter associated to class ω_q and α_0 is a fixed parameter. The BSs $\mathrm{m}^{s,1}, \mathrm{m}^{s,2}, \ldots, \mathrm{m}^{s,k}$ corresponding to the k nearest neighbors of $\boldsymbol{x}^s$ can then be combined using the Dempster's rule of combination [5]:

$$\mathrm{m} = \mathrm{m}^{s,1} \oplus \ldots \oplus \mathrm{m}^{s,k} \tag{4}$$

The focal elements of m are the classes of its k nearest neighbors, and Ω. The credibility and the plausibility of a given class ω_q are respectively:

$$\mathrm{bel}(\{\omega_q\}) = \mathrm{m}(\{\omega_q\}) \tag{5}$$

$$\mathrm{pl}(\{\omega_q\}) = \mathrm{m}(\{\omega_q\}) + \mathrm{m}(\Omega) \tag{6}$$

A probability distribution P such that $\mathrm{bel}(A) \leq P(A) \leq \mathrm{pl}(A)$ for all subset A of Ω is said to be *compatible* with m [5]. An example of such a probability distribution is the *pignistic* probability distribution BetP defined for each ω_q as [6]:

$$\mathrm{BetP}(\{\omega_q\}) = \mathrm{m}(\{\omega_q\}) + \frac{\mathrm{m}(\Omega)}{M}. \tag{7}$$

A decision regarding the classification of $\boldsymbol{x}^s$ can be made by assigning it to the class ω_{qmax} with maximum credibility (or equivalently, with maximum plausibility or pignistic probability):

$$qmax = \arg\max_q \mathrm{m}(\{\omega_q\}) \tag{8}$$

The parameters γ_q used to define the basic probability assignment are learnt from the training set by minimizing the mean squared error between the classifier's outputs and target values [10].

3 Generalization based on possibility theory

3.1 Belief structures and possibility measures

The above method may be generalized by exploiting certain relationships between the concepts of fuzzy set, possibility distribution and belief structure. As remarked by Dubois and Prade [2], a possibility measure is a special kind of plausibility function whose focal elements are nested. The corresponding necessity measure is the associated belief function, which is said to be *consonant*. In [9], the theory of possibility is also related to the theory of fuzzy sets by defining a possibility distribution as a membership

[1]The more general case in which the class membership of training patterns is itself subject to uncertainty has also been studied in [1].

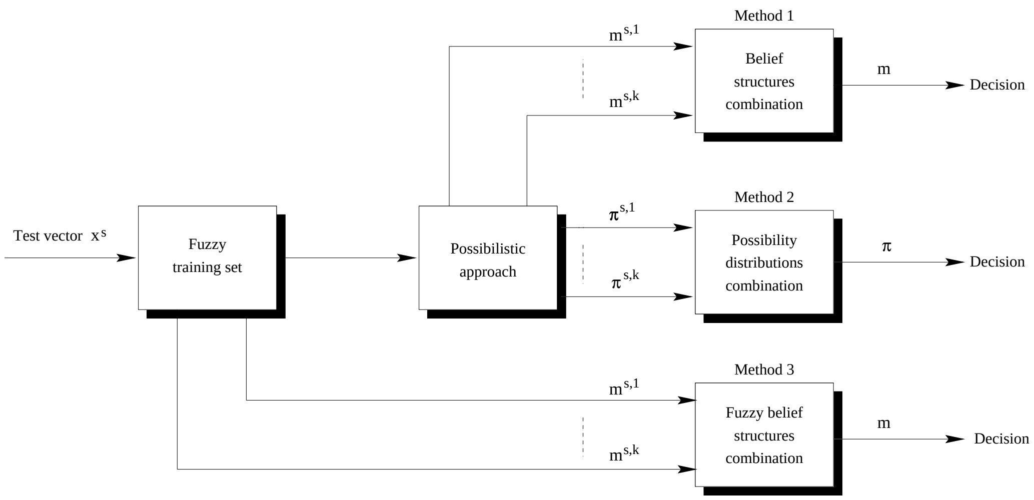

Figure 1: Illustration of the proposed methods.

function of a fuzzy set. More specifically, let F be a fuzzy subset of Ω characterized by its membership function μ_F. The fuzzy set F induces a possibility distribution π equal to the membership function of F:

$$\pi(\omega) = \mu_F(\omega) \quad \forall \omega \in \Omega \tag{9}$$

The possibility and necessity measures induced by π are defined as:

$$\Pi(A) = \max_{\omega \in A} \pi(\omega) \tag{10}$$

$$N(A) = \min_{\omega \in A} (1 - \pi(\omega)) \tag{11}$$

for all $A \subseteq \Omega$. Functions Π and N are respectively a credibility and a plausibility measure, whose focal elements are the α-cuts of F. Furthermore, let $\pi_1 = 1 > \pi_2 > \ldots > \pi_r$ be the set of distinct values of π, F_i the π_i-cut of F $(i = 1, \ldots, r)$, and $\pi_{r+1} = 0$ by convention. The belief structure m underlying Π is then completely defined in terms of the possibility distribution π as:

$$\mathrm{m}(F_i) = \pi_i - \pi_{i+1} \quad \forall i = 1, \ldots, r \tag{12}$$

The link between belief functions and possibility distributions allows to consider a variety of combination methods. In particular, the combination of two possibility distributions π_1 and π_2 can be defined by means of fuzzy set-theoretic operations:

$$\pi(\omega) = \frac{\pi_1(\omega) * \pi_2(\omega)}{\max_{\omega \in \Omega} \ \pi_1(\omega) * \pi_2(\omega)} \tag{13}$$

where $*$ is a triangular norm such as the min operator. The result of this combination produce a possibility distribution. Alternatively, the combination can also be performed using Dempster's rule. However, the BS resulting from the combination of two consonant BSs m^1 and m^2 associated to possibility distributions π_1 and π_2 is generally not consonant, and therefore does not correspond to a possibility distribution.

3.2 Application to Pattern Recognition

In the following, we assume each training pattern $\boldsymbol{x}^i$ to be labeled by a membership vector $\mu_i = (\mu_1^i, \ldots, \mu_M^i)^t$, where μ_j^i denotes the membership degree of pattern $\boldsymbol{x}^i$ to fuzzy class ω_j. We can therefore associate to $\boldsymbol{x}^i$ a fuzzy subset F^i of Ω as follows:

$$F^i = \frac{\mu_1^i}{\omega_1} + \ldots + \frac{\mu_M^i}{\omega_M} \tag{14}$$

As explained above, this fuzzy set may be converted into a consonant belief structure m^i.

Given a new pattern $\boldsymbol{x}^s$ to be classified, the information provided by training pattern $\boldsymbol{x}^i$ and its class label may be represented by a BS $m^{s,i}$ based on m^i and the distance $d^{s,i}$ between $\boldsymbol{x}^i$ and $\boldsymbol{x}^s$:

$$\begin{aligned} m^{s,i}(A) &= \alpha^{s,i} m^i(A) \quad \forall A \subseteq \Omega && (15)\\ m^{s,i}(\Omega) &= 1 - \alpha^{s,i} \sum_{A \neq \Omega} m^i(A) && (16) \end{aligned}$$

where $\alpha^{s,i}$ is defined as a decreasing function of $d^{s,i}$:

$$\alpha^{s,i} = \alpha_0 \exp(-\gamma (d^{s,i})^2) \quad (17)$$

The focal elements of $m^{s,i}$ are the π_i-cut of F^i.

Two methods can be considered for pooling the evidence of the k-NN of $\boldsymbol{x}^s$ (Fig. 1):

1. combination of the BSs $m^{s,i}$ corresponding to the k-NN of $\boldsymbol{x}^s$ using Dempster's rule (Method 1);
2. combination of the possibility distributions $\pi^{s,i}$, associated to the BSs $m^{s,i}$ using a possibilistic rule, the resulting possibility distribution π inducing a consonant belief structure (Method 2).

The parameter γ appearing in Eq. 17 can be determined by minimizing an error criterion computed on the training set. According to the method used, we propose two criteria $E_1(\boldsymbol{x})$ (method 1) and $E_2(\boldsymbol{x})$ (method 2) defined respectively as:

$$\begin{aligned} E_1(\boldsymbol{x}) &= \frac{1}{2}\sum_{j=1}^{M} (\mathrm{BetP}(\{\omega_j\}) - \mu_j)^2 && (18)\\ E_2(\boldsymbol{x}) &= \frac{1}{2}\sum_{j=1}^{M} (\pi(\omega_j) - \mu_j)^2 && (19) \end{aligned}$$

where μ_j is the class membership of training pattern $\boldsymbol{x}$ to fuzzy class ω_j, and BetP is the pignistic probability distribution corresponding to the BS m obtained from the k-NN of $\boldsymbol{x}$. Criterion E_1 generalizes the error function proposed in [10, 11] in the case of crisp classes.

4 Generalization based on fuzzy belief structures

4.1 Generalization of the D-S theory to fuzzy sets

Shafer's theory was developed under the condition that each BS m is defined over crisp subsets of Ω. Several attempts have been made to extend this theory to fuzzy sets. We consider here the approach proposed by Yager [7, 8]. Let $\mathcal{F}$ be the set of all fuzzy subsets of Ω. A *fuzzy belief structure* is a function $m : \mathcal{F} \longmapsto [0,1]$ defined as follow:

1. The set $\mathcal{F}'$ containing those $A \in \mathcal{F}$ such that $m(A) > 0$ is finite. The elements of $\mathcal{F}'$ are called the *focal elements* of the fuzzy belief structure.
2. $\sum_{A \in \mathcal{F}'} m(A) = 1$.
3. $\forall A \in \mathcal{F}, \max_{\omega \in \Omega} \mu_A(\omega) < 1 \Rightarrow m(A) = 0$.

A function m that respects conditions 1 and 2, but not condition 3 is called a *pseudo-belief structure* by Yager. The plausibility and credibility of any $A \in \mathcal{F}$ can then be defined by:

$$\begin{aligned} \mathrm{bel}(A) &= \sum_{B \in \mathcal{F}'} m(B)\mathrm{Inc}(B, A) && (20)\\ \mathrm{pl}(A) &= \sum_{B \in \mathcal{F}'} m(B)\mathrm{Int}(A, B) && (21) \end{aligned}$$

where $\mathrm{Int}(B, A)$ is the degree of intersection between the two fuzzy subsets A and B, and $\mathrm{Inc}(B, A)$ is the degree of inclusion of B in A, defined respectively as:

$$\begin{aligned} \mathrm{Int}(A, B) &= \max_{\omega \in \Omega}[\min(\mu_A(\omega), \mu_B(\omega))] && (22)\\ \mathrm{Inc}(B, A) &= \min_{\omega \in \Omega}[\max(\mu_A(\omega), \mu_{\bar{B}}(\omega))] && (23) \end{aligned}$$

where $\mu_{\bar{B}}(\omega) = 1 - \mu_B(\omega)$. A generalization of Dempster's rule of combination has also been proposed by Yager. This rule consists in two operations. Fuzzy BSs are first combined using a generalization of the Dempster's rule without normalization:

$$m(A) = \sum_{B, C \in \mathcal{F}', B \cap C = A} m^1(B) m^2(C) \quad \forall A \in \mathcal{F} \quad (24)$$

The function m is generally a pseudo-structure whose focal elements F^i may be subnormal fuzzy subsets of $\mathcal{F}$. The normalized combined BS m^* can be obtained from m by using the following normalization process [8]:

1. Introduce a new fuzzy subsets E_i with membership degree $\mu_{E_i}(\omega) = \mu_{F_i}(\omega)\ /h_i$ and $h_i = \max_{\omega \in \Omega} \mu_{F_i}(\omega)$.
2. Note $T = \sum_i m(F_i)(1 - h_i)$.

3. Compute $\mathrm{m}^*(E_i) = \sum_i \mathrm{m}(F_i)h_i \ /(1-T)$

Some properties of m^* are given in [8]. If we need to combine more than two BSs $\mathrm{m}^1, \ldots, \mathrm{m}^k$, the result of the combination is computed by first combining the BS without normalization using (Eq. 24), and then normalizing the final result.

4.2 Application to Pattern Recognition

Let us consider a pattern $\boldsymbol{x}^s$ to be classified. We assume that each training pattern $\boldsymbol{x}^i$ is labeled by a fuzzy subset F^i of Ω, as described previously. For each neighbor $\boldsymbol{x}^i$ of $\boldsymbol{x}^s$, a fuzzy BS $\mathrm{m}^{s,i} : \mathcal{F} \mapsto [0,1]$ is defined as follows:

$$\mathrm{m}^{s,i}(F^i) = \alpha^{s,i} \tag{25}$$

$$\mathrm{m}^{s,i}(\Omega) = 1 - \alpha^{s,i} \tag{26}$$

and $\mathrm{m}^{s,i}(A) = 0$ for all $A \in \mathcal{F} \setminus \{\Omega, F^i\}$. As in the original approach, $\alpha^{s,i}$ is a decreasing function of the distance between $\boldsymbol{x}^s$ and $\boldsymbol{x}^i$. A fuzzy BS m associated to $\boldsymbol{x}^s$ is determined by combining the BSs $\mathrm{m}^{s,i}$ using the generalized Dempster's rule. The credibility and plausibility of each element of $\mathcal{F}$ can then be computed using Eq. 20 and 21. A decision regarding the classification of $\boldsymbol{x}^s$ can be made by assigning it to the class with maximum credibility or maximum plausibility. The method used to optimize the fuzzy BS assigned to $\boldsymbol{x}^s$ generalizes the method originally proposed in [10]: we attempt to minimize the mean squared error between the fuzzy membership vector $\mu = (\mu_1, \ldots, \mu_M)^t$ of each training pattern $\boldsymbol{x}$, and the vector $(\mathrm{pl}(\{\omega_1\}, \ldots, \mathrm{pl}(\{\omega_M\}))^t$ of plausibilities computed using the neighbors of $\boldsymbol{x}$ in the training set:

$$E(\boldsymbol{x}) = \frac{1}{2}\sum_{j=1}^{M}(\mathrm{pl}(\{\omega_j\}) - \mu_j)^2 \tag{27}$$

The parameter γ is then determined by minimizing the mean error over the training set.

5 Results

The above methods were tested using two artificial classification tasks. The first task concerns three non Gaussian classes separated by a non linear boundary (Fig. 2). The second tasks involves three Gaussian classes with the following characteristics:

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

For each problem, a training set of 90 samples and a test set of 600 samples were generated. For each training pattern $\boldsymbol{x}$, a membership degree μ_i with respect to fuzzy class ω_i was defined as:

$$\mu_i = \frac{1}{1 + d(\boldsymbol{x}, \bar{\boldsymbol{x}}^i_k)} \tag{28}$$

where $\bar{\boldsymbol{x}}^i_k$ represents the mean vector of the k-NN of $\boldsymbol{x}$ belonging to class ω_i. This method allows to fuzzify the class labels of training patterns using local information.

The performance of the proposed methods were compared with those of the voting k-NN, the evidence theoretic k-NN and the fuzzy k-NN rules introduced by Keller [4]. The Euclidean metric was used as a distance measure. The tie problem in the voting k-NN rule was resolved randomly. The parameters of the fuzzy k-NN rule were chosen as recommended by Keller.

As shown in Fig. 3, the three methods described in this paper achieved the same performance on the two problems, and significantly outperformed the other methods. These results demonstrate the usefulness of the proposed methods for taking into account fuzziness in the classification procedure.

6 Conclusion

In this paper, two main approaches for processing training data with fuzzy class labels in the framework of evidence-theoretic pattern recognition have been described. The first approach exploits the relationship between fuzzy sets, possibility distributions and consonant belief structures. The evidence of the k nearest neighbors of a pattern to be classified is then pooled using either Dempster's rule of combination or a possibilistic rule such as fuzzy set intersection. The second approach is based on a generalization of the Dempster-Shafer theory to fuzzy sets. It consists in representing the evidence of each training pattern by a belief structure with fuzzy focal elements. The

(a) (b)

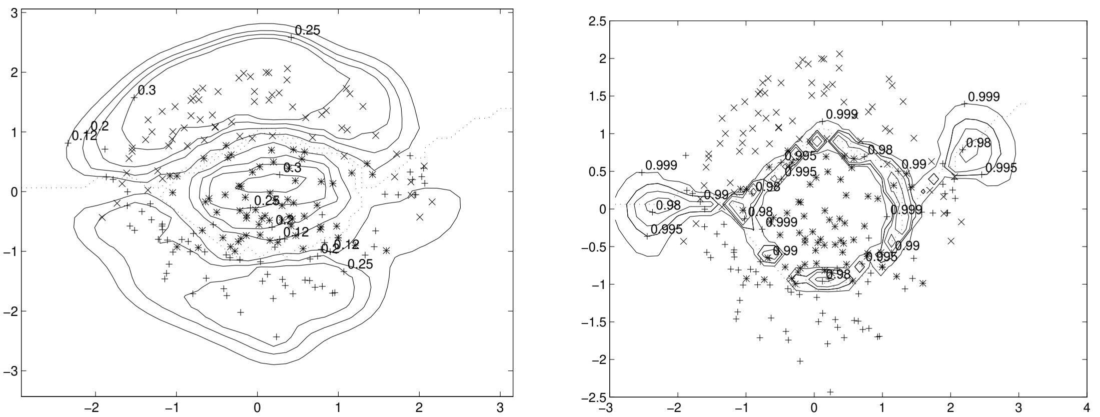

Figure 2: Contour lines of the credibility (a) and plausibility (b) functions with fuzzy focal elements (obtained by method 3).

(a) (b)

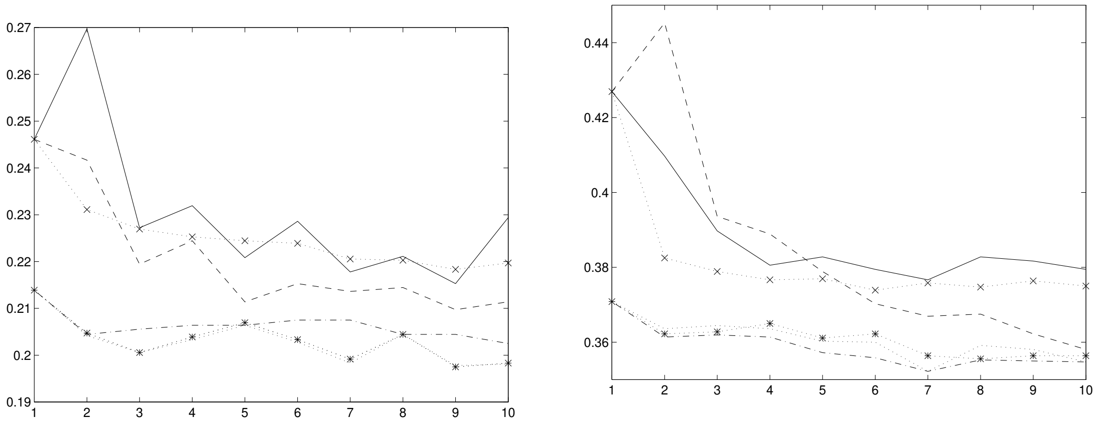

Figure 3: Test error rates on data sets: (a) non Gaussian data, (b) Gaussian data, for k ranging from 1 to 10. The methods are the voting k-NN rule (-), the evidence-theoretic k-NN (- -), the fuzzy k-NN ($.\times$), the k-NN method 1 (..), the k-NN method 2 ($.*$) and the k-NN method 3 (-.).

combination of the different structures is then performed using the generalized Dempster rule. The proposed schemes were shown experimentally to be roughly equivalent in terms of error rate, and to outperform other methods using the same information. The method based on fuzzy belief structures seems to be the most general of all three, as it can be further generalized to allow class labels to be represented by fuzzy belief structures.

References

[1] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(05):804–813, 1995.

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

[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*, 12:193–226, 1986.

[4] J. M. Keller, M. R. Gray, and J. A. Givens. A fuzzy k-NN neighbor algorithm. *IEEE Trans. Syst. Man Cybern.*, SMC-15(4):580–585, 1985.

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

[6] P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[7] R. R. Yager. Generalized probabilities of fuzzy events from fuzzy belief structures. *Information Sciences*, 28:45–62, 1982.

[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(8):1221–1230, 1995.

[9] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.

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

[11] L. M. Zouhal and T. Denœux. A comparison between fuzzy and evidence-theoretic k-NN rules for pattern recognition. In *Proceedings of EUFIT'95*, volume 3, pages 1319–1325, Aachen, August 1995.