New neighborhood classifiers based on evidential reasoning

Deqiang Han
Inst. of Integrated Automation
MOE KLINNS Lab
Xi'an Jiaotong University
Xi'an, China 710049
deqhan@gmail.com

Jean Dezert
The French Aerospace Lab
ONERA
Chemin de la Hunière
F-91761 Palaiseau, France
jean.dezert@onera.fr

Yi Yang
SKLSVMS
School of Aerospace
Xian Jiaotong University
Xian, China 710049
jiafeiyy@mail.xjtu.edu.cn

Chongzhao Han
Inst. of Integrated Automation
MOE KLINNS Lab
Xi'an Jiaotong University
Xi'an, China 710049
czhan@mail.xjtu.edu.cn

***Abstract*—Neighborhood based classifiers are commonly used in the applications of pattern classification. However, in the implementation of neighborhood based classifiers, there always exist the problems of uncertainty. For example, when one use *k*-NN classifier, the parameter *k* should be determined, which can be big or small. Therefore, uncertainty problem occurs for the classification caused by the *k* value. Furthermore, for the nearest neighbor (NN) classifier, one can use the nearest neighbor or the nearest centroid of all the classes, so different classification results can be obtained. This is a type of uncertainty caused by the local and global information used, respectively. In this paper, we use theory of belief function to model and manage the two types of uncertainty above. Evidential reasoning based neighborhood classifiers are proposed. It can be experimentally verified that our proposed approach can deal efficiently with the uncertainty in neighborhood classifiers.**

Keywords—neighborhood classifier; uncertainty; belief functions; evidential reasoning; information fusion.

I. INTRODUCTION

Neighborhood classifiers, such as the most commonly used nearest neighbor (NN) introduced by Fix and Hodges [1] and k-nearest neighbor [2], etc. are widely discussed and applied in pattern classification and machine learning. A wide variety of neighborhood classifiers have been developed [3], [4], [5], [6].

In practical use of the neighborhood classifiers, there always exist the problems of uncertainty. For example, the value of the parameter k that controls the size of the neighborhood has to be determined. The selection of k is always ad hoc and depends on the user's preference. Different k values (small, moderate or big) correspond to different size of neighborhood, which generally yields different results. Although one can use the cross-validation approach [2] to determine an "optimal" k value by minimizing the overall probability of error, the optimal value of k at one point does not have to be, and in general is not, the same as the optimal value of k at some other point. It should be also noted that cross-validation will require a large computational cost. Some researchers designed adaptive k value selection approaches. For example, Wang et al. [7] proposed a locally adaptive selection of the k by using their defined statistical confidence. Ghoshet al. [8] presented an adaptive nearest neighbor classification technique, where the value of k is selected depending on the distribution of competing classes in the vicinity of the observation to be classified. Other researchers proposed some non-parametric k-NN where the k is automatically determined by geometric relationships. See [9] for details. Furthermore, in the nearest neighbor (NN) classifier, one can use the nearest neighbor (local information) to make the classification. One can also use the nearest centroid [10] of all the classes (they can relatively globally represent the classes) to make the classification. So, different classification results can be obtained. It should be better if we can jointly use the two types of "NN" to deal with the information provided by the local (nearest neighbor) and global information (nearest centroid). Such a joint use of global and local information requires addressing the problem of uncertainty.

As aforementioned, for the neighborhood classifiers there are some types of uncertainty problems. The theory of belief function [11] is widely used in uncertainty management and uncertainty reasoning for decision-making. In this paper, we attempt to use it to model and manage the uncertainty incorporated in neighborhood classifiers.

For the choice of k for k-NN, we set two values, i.e., a big one and a small one, to constitute an interval of k value. The small one (corresponding to a small neighborhood therefore representing relatively local information) and the big one (corresponding to a big neighborhood therefore representing relatively global information) are used to execute the k-NN classification, respectively. Two different classification results (including the class label and each class's frequency of appearance within the neighborhood) always can be obtained. By using the different classification results, we can construct the matrix of expected possibility interval and generate bba's (basic belief assignments). Then based on Dempster's rule of combination, or other types of combination rules and some probabilistic transformation [12], the final decision can be made. The joint use of two values of parameter k of the k-NN approach that we propose avoids the difficult search for an optimal value of k. As it will be shown in the sequel, the joint use of local and global information as proposed in our new evidential reasoning based k-NN classifier (ER-k-NN) still performs better.

For the joint use of nearest neighbor (NN) and nearest class centroid (NC), we use the NN and NC, respectively to obtain different classification results (including the class label and their corresponding possibility generated by using the distance). We construct the matrix of expected possibility based on the different classification results and generate bba's. By using some combination rule and probabilistic transformation, we can obtain the final decision result. We name this new classifier as evidential reasoning based nearest neighbor and nearest centroid classifier (ER-NN-NC).

Experimental results based on some artificial datasets and some datasets in UCI [13] show that our evidential reasoning based neighborhood classifiers including ER-k-NN and ER-NN-NC, are effective and they can well manage the uncertainty incorporated in.

II. Uncertainty in Nearest Neighbors Classifiers

A. The uncertainty caused by the choice of k in k-NN

The k-Nearest Neighbor (k-NN) method classifies an observation $\boldsymbol{x}$ to the class, which is the most frequent in the neighborhood of $\boldsymbol{x}$. The size of this neighborhood is usually determined by a predefined parameter k. k-NN can be used to obtain good estimates of Bayes error because the k-NN probability of error asymptotically approaches Bayes error as proved in [2].

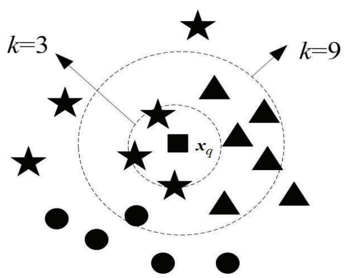

Fig. 1. Classification based on different k values in k-NN.

Let's consider the three classes problem as illustrated in Fig. 1. The three classes are labeled as Star (S), Triangle (T) and Circle (C) for convenience. The query sample x_q is represented by the black square. When k=3, there are three Star samples in the neighborhood of x_q, therefore, x_q will be labeled as belonging to the class Star. When $k = 9$, there are three Star samples, five Triangle samples, and one Circle sample in the neighborhood of x_q, therefore, x_q will be labeled as belonging to the class Triangle. As we can easily see in Fig. 1, different values of k bring different results. This is a well-known serious drawback of the k-NN classifier which reflects the impact of the uncertainty in the choice of k parameter. What criterion should one choose to determine the optimal value of k?

Cover and Hart [14] suggest that the value of k should depend on the training sample size n, and it should vary with n in such a way that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. However, when the sample size is small or moderately large, there is no theoretical guideline for choosing the value of k. The optimum value of k depends on the specific dataset, and one normally uses the usual cross-validation technique to estimate it. However, the computational cost of cross-validation is large, or prohibitive in some cases.

Another choice of k has been proposed by Wang et al. [7] based on a local adaptive selection strategy using statistical confidence.Ghosh [8] presented an adaptive nearest neighbor classification technique, where the value of k is selected based on the distribution of competing classes in the vicinity of the observation to be classified. Several researchers proposed also some non-parametric k-NN, such as relative graph neighbor (RGN) and Gabriel neighbor (GN)[9], where the k is automatically determined by geometric relationships. See the related references for details.

B. The uncertainty caused by the joint use of Nearest Neighbor (NN) and Nearest Class Centroid (NC)

The nearest neighbor (NN) classification is one of the simplest and popular methods for statistical pattern recognition. In fact NN is a non-parametric classifier [15]. In traditional NN classifier, only the local information is used. A given sample x_q which is most similar to one sample in a class c does not definitely represent that it is similar to the whole class c, therefore, x_q does not definitely belong to class c. The centroid of each class is more appropriate to represent the whole class when compared with a single sample. So, some researchers proposed a nearest class centroid classifier, where the distance between a given sample $x_q$and each class's centroid is calculated. The shortest distance's corresponding class is assigned to x_q. The difference of definition of the "nearest neighbor" can cause the different classification results as illustrated in Fig. 2.

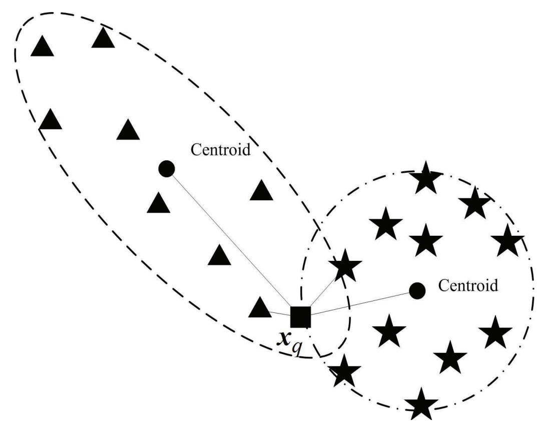

Fig. 2. Classification based on NN and NC.

In Fig. 2, for the given query sample x_q (square), it will be labeled as class Triangle according to NN while it will be labeled as class Star according to NC. Such a type of uncertainty of classification is caused by the different definitions of "nearest neighbor".

It should be noted that for the cases pointed out in this section, although the classification results could be the same according to different k or different definitions of NN, they can be different. More precisely, it can happen that k-NN classifiers with different k values can bring the same results for some samples and different results for other samples. This shows the inconstancy of k-NN due to the underlying uncertainty of the choice of k parameter. Then, how to deal with the uncertainty in such neighborhood classifiers?

The theory of belief functions is a powerful tool for modeling and managing uncertainty and that is why we propose to use it to improve classification performances. The basics of this theory are introduced in the next section.

III. Basics of the Theory of Belief Functions

The theory of belief functions introduced by Shafer in [11] is a theory of uncertain reasoning that is designed to deal with the distinction between uncertainty and ignorance. In this theory, elements in the frame of discernment (FOD) Θ are mutually exclusive and exhaustive. The power-set (the set of all subsets of Θ) is denoted by 2^Θ. A basic belief assignment (bba, or a mass function) is a mapping $m(.) : 2^\Theta \mapsto [0,1]$ which satisfies:

$$m(\emptyset) = 0, \quad \text{and} \quad \sum_{X \in 2^\Theta} m(X) = 1 \tag{1}$$

The belief function $Bel(.)$ and the plausibility function $Pl(.)$ are defined for all $X \in 2^\Theta$, respectively by

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y) \tag{2}$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) \tag{3}$$

Dempster's rule of combination (denoted DS for short) is the emblematic (and controversial) rule of combination of bba's proposed in Dempster-Shafer Theory (DST) [11]. DS rule is defined as follows. Let's consider $n \geq 2$ independent mass functions $m_1(.)$, $m_2(.)$, $\ldots$, $m_n(.)$, the new combined evidence obtained with DS rule is given by (4)

$$m_{DS}(X) = \begin{cases} 0, \ \text{if } X = \emptyset \\ \dfrac{\sum\limits_{\cap X_i = X} \prod\limits_{1 \leq i \leq n} m_i(X_i)}{\sum\limits_{\cap X_i \neq \emptyset} \prod\limits_{1 \leq i \leq n} m_i(X_i)} \ , \quad \text{if } X \neq \emptyset. \end{cases} \tag{4}$$

In the late 1970s, Zadeh [16] presented an example for which Dempster's rule of combination produces results usually judged as unsatisfactory or counter-intuitive when the bodies of evidence to be combined are highly conflicting. More recently, Dezert et al. [17], [18] have presented new important examples where Dempster's rule also fails to provide satisfactory results even in low conflicting cases as well. Since more than three decades, many researchers dispute the validity of the theory of belief functions, more precisely the validity of Dempster's rule of combination, and many rules for combining bba's have bloomed in the literature since a decade. The most appealing one so far is the proportional conflict redistribution rule no. 5 (PCR5) proposed in [12] and defined for two[1] bba's by: $m_{PCR5}(\emptyset) = 0$ and $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{PCR5}(X) = m_{12}(X) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[\frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)}\right] \tag{5}$$

All sets involved in the formula are in canonical form. $m_{12}(X)$ corresponds to the conjunctive consensus, i.e:

$$m_{12}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1) m_2(X_2).$$

[1] for $n > 2$ see [12].

All denominators are different from zero. If a denominator is zero, that fraction is discarded.

After combining the bba's by a given fusion rule we obtain a new bba. To make a decision on an element of the FOD Θ, we use a transformation to approximate the new bba in a probability mass function (pmf). The pignistic probability transformation $BetP(\cdot)$ proposed by Smets [19] is often used, which is illustrated in (6):

$$BetP_m(C_i) = \sum\nolimits_{\{C_i\} \in A \subseteq \Theta} m(A)/|A| \ \ , \quad \forall A \subseteq \Theta \tag{6}$$

As we will show, the joint use of local and global information in neighborhood classifiers can be used as a multi-criteria decision making (MCDM) problem. In some previous works, we have designed evidential-reasoning based MCDM approaches, such as COWA-ER and fuzzy COWA-ER related approaches [20], [21]. In the sequel, we implement new neighborhood classifiers based on evidential reasoning inspired by the COWA-ER approach.

IV. New Neighborhood Classifiers based on Evidential Reasoning

A. Evidential Reasoning based on k-NN (ER-k-NN)

As already mentioned, the choice of k in k-NN is a serious problem and that is why many attempts have been proposed to select k based on different strategies including cross-validation, graph neighborhood, adaptive choice, etc. In fact there is so far no consensus on the best theoretical guideline for choosing the value of k because the "best" value of k depends on the specific dataset. It cannot be selected by a strategy chosen a priori. To circumvent this problem of uncertainty in the choice of k we propose to use a bounded interval $[k_s, k_b]$ of k because it is much easier, and we expect to get a more robust result than with selecting only a single value for k. From such interval, it is relatively simple to model and use belief functions to make the classification exploiting both local and global information in order to get better performances. The new ER-k-NN method consists in the following steps:

• **Step 1:** Execute k-NN by using the bounds of $[k_s, k_b]$.

For the k-NN classifier, set two values for k, i.e., k_s(smaller) and k_b (bigger). Execute the k-NN algorithms twice according to $k_s$and k_b, respectively.

Let's consider a FOD $\Theta = \{C_1, C_2, ..., C_M\}$ with $M > 1$ classes. For each query sample x_q, one finds its k_s neighbors and its k_b neighbors, respectively. Then, one calculates the ratio (considered as a type of possibility) of the samples number belonging to each class C_i, $i = 1, 2, \ldots, M$ as follows:

$$p_s(C_i) = \frac{k_s^i}{k_s} \quad \text{and} \quad p_b(C_i) = \frac{k_b^i}{k_b} \tag{7}$$

where k_s^i is the number of samples belonging to class C_i in k_s neighbors and k_b^i is the number of samples belonging to class C_i in k_b neighbors. Obviously, one has

$$k_s = \sum_{i=1}^{M} k_s^i \quad \text{and} \quad k_b = \sum_{i=1}^{M} k_b^i \tag{8}$$

• **Step 2**: Construct the expected possibility matrix (EPM).

For the query sample x_q for each class, we calculate

$$\begin{cases} e_{\min}(C_i) = \min\{p_s(C_i), p_b(C_i)\} \\ e_{\max}(C_i) = \max\{p_s(C_i), p_b(C_i)\} \end{cases} \quad (9)$$

Then we can construct the expected possibility interval as $[e_{\min}(C_i), e_{\max}(C_i)]$ and the expected possibility matrix as

$$E(C) = \begin{bmatrix} [e_{\min}(C_1), e_{\max}(C_1)] \\ [e_{\min}(C_2), e_{\max}(C_2)] \\ \vdots \\ [e_{\min}(C_M), e_{\max}(C_M)] \end{bmatrix} \quad (10)$$

We divide each bound of intervals by the max of the bounds to get a new normalized expected possibility matrix as

$$E^{norm}(C) = \begin{bmatrix} [e^{norm}_{\min}(C_1), e^{norm}_{\max}(C_1)] \\ [e^{norm}_{\min}(C_2), e^{norm}_{\max}(C_2)] \\ \vdots \\ [e^{norm}_{\min}(C_M), e^{norm}_{\max}(C_M)] \end{bmatrix} \quad (11)$$

where

$$\begin{cases} e^{norm}_{\min}(C_i) = e_{\min}(C_i)/\max(E[C]) \\ e^{norm}_{\max}(C_i) = e_{\max}(C_i)/\max(E[C]) \end{cases} \quad (12)$$

• **Step 3:** Generate bba's using EPM.

Generate M bba's using $E^{norm}(C)$ as follows:

$$\begin{cases} m_i(C_i) = e^{norm}_{\min}(C_i) \\ m_i(\overline{C_i}) = 1 - e^{norm}_{\max}(C_i) \\ m_i(C_i \cup \overline{C_i}) = e^{norm}_{\max}(C_i) - e^{norm}_{\min}(C_i) \end{cases} \quad (13)$$

In this generation of bba's, the uncertainty is represented by length of the interval $[e^{norm}_{\min}(C_i), e^{norm}_{\max}(C_i)]$ and corresponds to the imprecision of the variable (here the expected possibility) on which the belief function for C_i is defined.

• **Step 4:** Combination of bba's obtained in Step 3.

This operation is represented by

$$m(.) = [m_1 \oplus m_2 \oplus \cdots \oplus m_M](.) \quad (14)$$

where $\oplus$ denotes symbolically the fusion operator. The choice of the combination rule $\oplus$ is left to the user, but in the sequel we have used and tested the DS rule given in (4) and PCR5 rule given in (5) and compared their performances with respect to the classical approaches.

• **Step 5:** Assign the object x_q to a single class based on $m(.)$.

For this, we need to use some decision-making procedure. The most common one is to approximate the combined bba $m(.)$ defined on 2^Θ into a subjective probability measure $P(.)$ defined on Θ. Then, one takes as assignment solution the class corresponding to the max of $P(.)$. Unfortunately, there exist many solutions based on different justifications to make such approximation which yield different decision-making results. In the following we will use the two common ones: the pignistic probability $BetP(.)$ [19], and $DSmP(.)$ proposed by Dezert and Smarandache in [12]. The mathematical formulas of $BetP(.)$ and $DSmP(.)$ are given in the corresponding references and are not reported here due to space limitation restriction.

Example 1: Let's take the query sample x_q in Fig. 1 and the frame of the three classes $\Theta = \{C_1 = Star, C_2 = Triangle, C_3 = Circle\}$. We choose as parameter interval $[k_s, k_b] = [3, 9]$. The step 1 of ER-k-NN gives us

$$p_s(C_1) = 1, p_s(C_2) = 0, p_s(C_3) = 0,$$
$$p_b(C_1) = 1/3, p_b(C_2) = 5/9, p_b(C_3) = 1/9.$$

The EPM obtained in step 2 of ER-k-NN is

$$E(C) = \begin{bmatrix} [e_{\min}(C_1), e_{\max}(C_1)] \\ [e_{\min}(C_2), e_{\max}(C_2)] \\ [e_{\min}(C_3), e_{\max}(C_3)] \end{bmatrix} = \begin{bmatrix} [\frac{1}{3}, & 1] \\ [0, & \frac{5}{9}] \\ [0 & \frac{1}{9}] \end{bmatrix}$$

Since in this example, the maximum value of $E(C)$ is 1, the normalized EPM $E^{norm}(C)$ equals to $E(C)$. Based on $E^{norm}(C)$, we generate bba's as in step 3 of ER-k-NN:

$$\begin{cases} m_1(C_1) = e^{norm}_{\min}(C_1) = 1/3 \\ m_1(\overline{C_1}) = m(C_2 \cup C_3) = 1 - e^{norm}_{\max}(C_1) = 0 \\ m_1(C_1 \cup \overline{C_1}) = m_1(\Theta) = e^{norm}_{\max}(C_1) - e^{norm}_{\min}(C_1) = 2/3 \end{cases}$$

$$\begin{cases} m_2(C_2) = e^{norm}_{\min}(C_2) = 0 \\ m_2(\overline{C_2}) = m_2(C_1 \cup C_3) = 1 - e^{norm}_{\max}(C_2) = 4/9 \\ m_2(C_2 \cup \overline{C_2}) = m_2(\Theta) = e^{norm}_{\max}(C_2) - e^{norm}_{\min}(C_2) = 5/9 \end{cases}$$

$$\begin{cases} m_3(C_3) = e^{norm}_{\min}(C_3) = 0 \\ m_3(\overline{C_3}) = m_3(C_1 \cup C_2) = 1 - e^{norm}_{\max}(C_3) = 8/9 \\ m_3(C_3 \cup \overline{C_3}) = m_3(\Theta) = e^{norm}_{\max}(C_3) - e^{norm}_{\min}(C_3) = 1/9 \end{cases}$$

By choosing Dempster's rule of combination followed by $BetP(.)$ transformation (steps 4 and 5 of ER-k-NN), we obtain $m_{DS}(.)$ with

$$m_{DS}(C_1) = 0.5967, m_{DS}(C_1 \cup C_3) = 0.0329$$
$$m_{DS}(C_1 \cup C_2) = 0.3292, m_{DS}(C_1 \cup C_2 \cup C_3) = 0.0412$$

and

$$BetP(C_1) = 0.7915, BetP(C_2) = 0.1783, BetP(C_3) = 0.0302$$

Based on this result, the final assignment of x_q will be C_1, i.e., the class Star.

By choosing PCR5 rule of combination followed by $DSmP(.)$ transformation (steps 4 and 5 of ER-k-NN), we obtain $m_{PCR5}(.)$ with

$$m_{PCR5}(C_1) = 0.5967, m_{PCR5}(C_1 \cup C_3) = 0.0329$$
$$m_{PCR5}(C_1 \cup C_2) = 0.3292,$$
$$m_{PCR5}(C_1 \cup C_2 \cup C_3) = 0.0412$$

and

$$DSmP(C_1) = 0.9993, DSmP(C_2) = 0.0006,$$
$$DSmP(C_3) = 0.0001$$

Based on this result, the final assignment of x_q will be C_1, i.e., the class Star.

B. Evidential Reasoning with NN and NC (ER-NN-NC)

Nearest neighbor (NN) classification is one of the simplest and popular methods for statistical pattern recognition. Nearest class centroid (NC) is also a kind of special "nearest neighbor" classifier. As aforementioned, if we use different definitions of nearest "neighbor" (nearest sample or nearest class centroid), the classification can be different. Here we also design evidential reasoning NN to jointly use the two different types of NN and to deal with such uncertainty. The new ER-NN-NC method consists in the following steps:

• **Step 1:** Execute NN and NC separately.

Let's consider a FOD $\Theta = \{C_1, C_2, ..., C_M\}$ with $M > 1$ classes. For each query sample x_q, one find its nearest neighbor in each class and calculate its distance with respect to x_q denoted d_{C_i} as shown in Fig. 3.

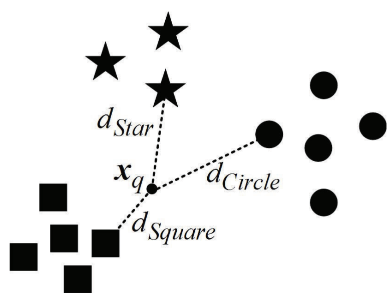

Fig. 3. Nearest neighbor in each class.

Based on these distances, one computes the probabilities to assign x_q with C_i by

$$P_{NN}(C_i) = \frac{e^{-d_{C_i}}}{\sum_{j=1}^{M} e^{-d_{C_j}}} \tag{15}$$

Similarly, one calculates the distance $d_{C_i}^{cen}$ between x_q and each class centroid as illustrated in Fig. 4.

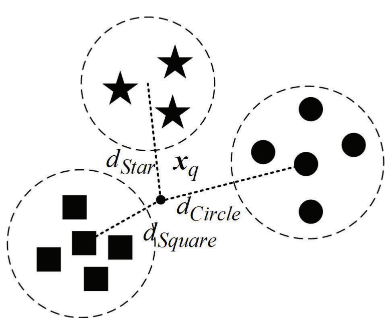

Fig. 4. Distance between query sample and each class centroid.

Based on these distances, one computes other probabilities to assign x_q with C_i by

$$P_{NC}(C_i) = \frac{e^{-d_{C_i}^{cen}}}{\sum_{j=1}^{M} e^{-d_{C_j}^{cen}}} \tag{16}$$

• **Step 2:** Construction of the EPM.

For the query sample x_q and for each class, we calculate

$$\begin{cases} e_{\min}(C_i) = \min\{P_{NN}(C_i), P_{NC}(C_i)\} \\ e_{\max}(C_i) = \max\{P_{NN}(C_i), P_{NC}(C_i)\} \end{cases} \tag{17}$$

Then we can construct the expected possibility interval as $[e_{\min}(C_i), e_{\max}(C_i)]$ and the EPM by (10). The normalized EPM $E^{norm}(C)$ is obtained by (11).

• **Steps 3–5:** bba's modeling, fusion and decision.

The steps 3–5 of ER-NN-NC method are the same as for ER-k-NN described in the previous subsection.

Example 2: Let's consider the frame of the three classes $\Theta = \{C_1, C_2, C_3\}$. For the query sample x_q, let's suppose that the distances to the nearest neighbors in each class are

$$d_{C_1} = 1,\ d_{C_2} = 1.5,\ d_{C_3} = 1.8$$

and the distances to each class centroid are

$$d_{C_1}^{cen} = 2,\ d_{C_2}^{cen} = 3.0,\ d_{C_3}^{cen} = 2.8$$

According to Step 1 of ER-NN-NC, one gets

$$p_{NN}(C_1) = 0.3185, p_{NN}(C_2) = 0.1172, p_{NN}(C_3) = 0.0643$$

$$p_{NC}(C_1) = 0.3440, p_{NC}(C_2) = 0.0466, p_{NC}(C_3) = 0.0695$$

The step 2 of ER-NN-NC provides the following EPM

$$E(C) = \begin{bmatrix} [e_{\min}(C_1), e_{\max}(C_1)] \\ [e_{\min}(C_2), e_{\max}(C_2)] \\ [e_{\min}(C_3), e_{\max}(C_3)] \end{bmatrix} = \begin{bmatrix} [0.3185, & 0.3440] \\ [0.0466, & 0.1172] \\ [0.0643 & 0.0695] \end{bmatrix}$$

having the maximum value 0.3440. The normalized EPM given by (11) is

$$E^{norm}(C) = \begin{bmatrix} [0.9259, & 1.0000] \\ [0.1353, & 0.3406] \\ [0.1869 & 0.2019] \end{bmatrix}.$$

Based on $E^{norm}(C)$, we generate bba's by (13) and one gets

$$\begin{cases} m_1(C_1) = e_{\min}^{norm}(C_1) = 0.9259 \\ m_1(\overline{C_1}) = m(C_2 \cup C_3) = 1 - e_{\max}^{norm}(C_1) = 0 \\ m_1(C_1 \cup \overline{C_1}) = m_1(\Theta) = e_{\max}^{norm}(C_1) - e_{\min}^{norm}(C_1) = 0.0741 \end{cases}$$

$$\begin{cases} m_2(C_2) = e_{\min}^{norm}(C_2) = 0.1353 \\ m_2(\overline{C_2}) = m_2(C_1 \cup C_3) = 1 - e_{\max}^{norm}(C_2) = 0.6594 \\ m_2(C_2 \cup \overline{C_2}) = m_2(\Theta) = e_{\max}^{norm}(C_2) - e_{\min}^{norm}(C_2) = 0.2053 \end{cases}$$

$$\begin{cases} m_3(C_3) = e_{\min}^{norm}(C_3) = 0.1869 \\ m_3(\overline{C_3}) = m_3(C_1 \cup C_2) = 1 - e_{\max}^{norm}(C_3) = 0.7981 \\ m_3(C_3 \cup \overline{C_3}) = m_3(\Theta) = e_{\max}^{norm}(C_3) - e_{\min}^{norm}(C_3) = 0.0150 \end{cases}$$

By choosing Dempster's rule of combination followed by $BetP(.)$ transformation (steps 4 and 5 of ER-NN-NC), we obtain

$$m_{DS}(C_1) = 0.9541, m_{DS}(C_2) = 0.0113,$$
$$m_{DS}(C_3) = 0.0166, m_{DS}(C_1 \cup C_3) = 0.0010,$$
$$m_{DS}(C_1 \cup C_2) = 0.0167, m_{DS}(C_1 \cup C_2 \cup C_3) = 0.0003.$$

and

$$BetP(C_1) = 0.9631, BetP(C_2) = 0.0198, BetP(C_3) = 0.0172$$

Based on this result, the final assignment of x_q is the class C_1.

By choosing PCR5 rule of combination followed by $DSmP(.)$ transformation (steps 4 and 5 of ER-k-NN), we obtain $m_{PCR5}(.)$

$$\begin{aligned} &m_{PCR5}(C_1) = 0.8433, m_{PCR5}(C_2) = 0.0188, \\ &m_{PCR5}(C_3) = 0.0334, m_{PCR5}(C_1 \cup C_3) = 0.0432 \\ &m_{PCR5}(C_1 \cup C_2) = 0.0551, m_{PCR5}(C_2 \cup C_3) = 0, \\ &m_{PCR5}(C_1 \cup C_2 \cup C_3) = 0.0061 \end{aligned}$$

and

$$\begin{aligned} &DSmP(C_1) = 0.9444, DSmP(C_2) = 0.0202, \\ &DSmP(C_3) = 0.0354. \end{aligned}$$

Based on this result, the final assignment of x_q will be C_1, i.e., the class Star.

V. Simulation and Experimental Results

In this section we present a comparative analysis of the performances of different classifiers based on an artificial dataset and on a real dataset.

A. Classification results based on artificial datasets

In this experiment, we generate 2D samples (x, y) belonging to a class C_1 according to the distribution (18)), and samples belonging to a class C_2 according to the distribution (19).

$$p_1(x,y) = \begin{cases} \frac{1}{b-a} \times \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{x}{\sigma})^2], & \text{if } y \in [a,b] \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

$$p_2(x,y) = \begin{cases} \frac{1}{b-a} \times \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2], & \text{if } y \in [a,b] \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The abscissa x follows a Gaussian distribution whereas the ordinate y follows a uniform distribution on $[a, b]$. In (19), μ denotes the average distance between the two centers of Gaussian pdf. Five hundred samples have been randomly generated (250 samples in class C_1 and 250 samples in class C_2). Randomly, we select 125 training samples from each class, the remaining samples are used for testing the performances of the classifiers. In this simulation, we have chosen $\mu = 1.5$, $\sigma = 1$ and $[a, b] = [-3, 3]$. The examples generated are shown in Fig. 5.

The evaluation of the performances of the different methods tested is based on a 100 runs Monte-Carlo simulation.

In each run, the samples selected for the training and the test datasets are drawn randomly. We have evaluated the performances (defined as the percentage of correct classification) of the nearest neighbor (NN) classifier, the nearest class centroid (NC) classifier, two classical k-NN classifiers (one is with a big k value and the other with a small k value), the ER-k-NN classifier, and the ER-NN-NC classifier. The results obtained are listed in Table I. To make comparisons, we also have done the cross-validation to find the optimal k for this experiment.

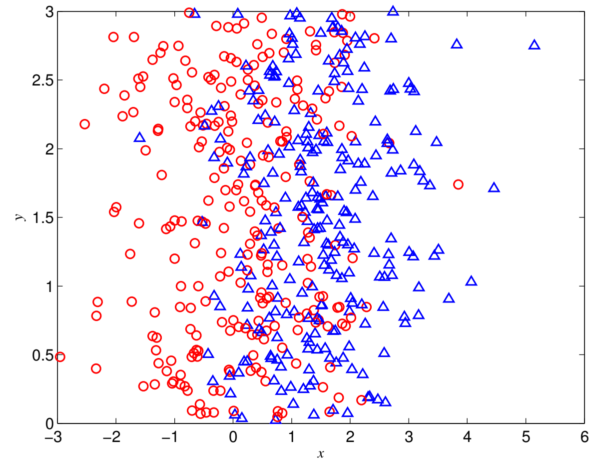

Fig. 5. Samples with a compound distribution.

TABLE I. Classification performance on Gaussian distributed datasets

Classification approaches	Averaged Classification Accuracy (%)
NN	73.51
NC	78.72
ER-NN-NC (DS+BetP)	**79.39**
ER-NN-NC (PCR5+DSmP)	**79.39**
k-NN (Big, k=120)	78.89
k-NN (Small, k=5)	76.08
k-NN (Optimal, $k = 115$)	**79.61**
ER-k-NN (DS+BetP)	77.81
ER-k-NN (PCR5+DSmP)	77.81

As we can see in TABLE I, the classification performances of ER-NN-NC (DS+BetP) and ER-NN-NC (PCR5+DSmP) are higher than those of NN and NC. That is to say, using both the local and global information, one improves significantly the classification results. The performances of ER-NN-NC (DS+BetP) and those of ER-NN-NC (PCR5+DSmP) are the same in this experiment because there is no special case where Dempter's rule performs counter-intuitively as those reported in [17], [18]. The classification performance of ER-k-NN is better than with the k-NN (k=5). Although it is not the best in this experiment, it is also not the worst one. The optimal choice of k for k-NN brings the best performance, however it requires a large computational cost of cross-validation which can become not acceptable for some applications manipulating a huge amount of data.

B. Classification results based on real UCI datasets

We have also tested the different classifiers on real datasets given in the machine learning repository of the University of California Irvine (UCI) [13] and listed in Table II,

In our tests, we do not deal with the missing data problem, all the samples with missing values have been eliminated. Features of the samples are normalized by their means and standard deviations before their classification.

As with the artificial datasets, we have evaluated the nearest neighbor (NN) classifier, the nearest class centroid (NC) classifier, two k-NN classifiers (one is with big k and the other with a small k), the ER-k-NN, and the ER-NN-NC classifier (both with DS+BetP option, and with PRC5+DSmP option). To make comparisons, we also have done the cross-validation to find the optimal k on each dataset used in this experiment. The results are listed in TABLE III.

TABLE II. UCI DATASETS USED IN THE EXPERIMENTS

Datasets	Class Num.	Feature dimension	Sample Num.
Iris	3	4	150
Wine	3	13	178
Pima	2	8	768
Bupa	2	6	345
Ionosphere	2	34	351

TABLE III. CLASSIFICATION ACCURACIES ON UCI DATASETS

Classifiers	Iris(%)	Wine(%)	Pima(%)	Bupa(%)	Iono(%)
NN	93.84	94.76	69.04	60.46	84.41
NN(Center)	92.09	95.68	72.70	56.54	79.25
ER-NN-NC (DS+BetP)	**95.15**	**96.42**	**73.38**	**60.96**	**87.76**
ER-NN-NC (PCR5+DSmP)	**95.16**	**96.44**	**73.38**	**60.96**	**87.76**
k-NN (k=40)	89.43	95.28	71.60	61.99	67.63
k-NN (k=5)	95.65	95.88	72.10	59.57	82.41
k-NN (Optimal k)	**96.04** (k = 11)	**96.78** (k = 9)	**74.37** (k = 33)	**62.81** (k = 23)	**84.22** (k = 1)
ER_k-NN(DS+BetP)	**95.85**	**96.30**	**72.90**	**62.35**	78.76
ER-k-NN (PCR5+DSmP)	**95.85**	**96.30**	**72.90**	**62.35**	78.76

As we can see in TABLE III, the ER-NN-NC (DS+BetP) method always performs better. ER-k-NN also performs well and provides a classification performance close to the best one obtained with optimal k value. This is due to the joint use of local and global information.

It should be noted that the classification performance of ER-k-NN or ER-NN-NC are not always the best for all kinds of datasets, but they perform well and their implementation is quite simple. The choice of interval $[k_s, k_b]$ in ER-k-NN is much easier than the search of the optimal value of k of k-NN. By using these ER based classifiers, the uncertainty is well modeled and managed.

Our previous research works [12] in information fusion have shown that PCR5 rule outperforms Dempster's rule specially when sources are highly conflicting, and in some emblematic examples with low conflicting sources as well [17], [18]. We have also shown that DSmP outperforms BetP in term of probabilistic information content [12]. These theoretical advantages of "PCR5 + DSmP" approach over "DS+BetP" approach have not been observed in our tests because our proposed ER-based neighborhood classifiers are some types of the multiple classifier ensemble, where the diversity among different member classifiers is the most crucial to improve ensemble classification performance as proved by Kuncheva and Whitaker in [22]. The choice of the fusion rule has a relative minor impact in fact. The chance to fall in a very specific "pathological" case for Dempster's rule is very small in our pattern classification experiments here, and this explains why "PCR5+DSmP" brings not much difference in term of performance with respect to "DS+BetP" in this work.

VI. CONCLUSIONS

In this paper, two kinds of ER-based neighborhood classifiers have been proposed to manage uncertainties by using belief function theory and the COWA-ER approach. Our simulation results show clearly that these new ER-based neighborhood classifiers perform well and their implementation are relatively simple since the selection of parameters is simplified. It should be noted that when the number of values for k is greater than 2, COWA-ER or Fuzzy COWA-ER can also be used to deal with the uncertainty to construct ER-based neighborhood classifiers.

In the implementation of these ER-based neighborhood classifiers, different types of combination rules and probabilistic transformations can be used which give some flexibility to the users. In this paper, only two combination rules (Dempster's rule and PCR5) and two probabilistic transformations (BetP and DSmP) have been tested. Of course many more could be implemented and tested as well and this is left for future investigations. The selection of the interval used in ER-k-NN classifier is an open question and we plan to make investigations on this question, and evaluate the robustness of ER-k-NN in future research works.

ACKNOWLEDGMENT

This study was co-supported by Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation of China (No.61104214, No. 61203222), Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 61221063), China Postdoctoral Science Foundation (No. 20100481337), China Postdoctoral Science Foundation - Special fund (No. 201104670), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20120201120036), and Fundamental Research Funds for the Central Universities (No. xjj2012104).

REFERENCES

[1] E. Fix, J. Hodges, Discriminatory analysis, nonparametric discrimination: consistency properties. Tech. Rep. 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

[2] O. R. Duda, E.P. Hart, D.G. Stork. Pattern Classification. New York: Wiley Inter-Science Publication, 2001.

[3] C. Zhou, Y. Yan, & Q. Chen. "Improving nearest neighbor classification with cam weighted distance," Pattern Recognition, 2006, vol. 39, p. 635–645.

[4] K. I. Kuncheva, C.J. Lakhmi. "Nearest neighbor classifier: simultaneous editing and feature selection," Pattern Recognition Letters, 1999, vol. 20, p. 1149–1156.

[5] M. Lindenbaum, S. Markovitch, D. Rusakov. "Selective sampling for nearest neighbor classifiers," Machine Learning, 2004, 54, p. 125–152.

[6] E. Kushilevitz, R. Ostrovsky, Y. Rabani. "Efficient search for approximate nearest neighbor in high dimensional spaces," SIAM Journal on Computing, 2000, 30, p. 457–474.

[7] J.G. Wang, P. Neskovic, L. N. Cooper. "Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence" Pattern Recognition, 2006, vol. 39, p. 417-423.

[8] A. K. Ghosh "On nearest neighbor classification using adaptive choice of k," Journal of computational and graphical statistics, 2007, vol. 16, no. 2, p. 271–350.

[9] J. S. Sanchez, F. Pla, F. J. Ferri. "On the use of neighbourhood-based non-parametric classifiers, " Pattern Recognition Letters, 1997, vol. 18, p. 1179-1186.

[10] B. B. Chaudhuri. "A new definition of neighbourhood of a point in multi-dimensional space, " Pattern Recognition Letters, 1996, vol. 17, p. 11–17.

[11] G. Shafer. A Mathematical Theory of Evidence, Princeton, NJ: Princeton University Press, 1976.

[12] F. Smarandache, J. Dezert (Editors). Applications and Advances of DSmT for Information Fusion (Vol III). Rehoboth, NM, USA: American Research Press, 2009. See fs.gallup.unm.edu/DSmT.htm

[13] C. L. Blake, C. L. Merz. UCI repository of machine learning databases. 1998. See http://www.ics.uci.edu/∼mlearn/MLRepository.html

[14] T. M. Cover, P. E., Hart. "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, 1967, vol. 13, p. 21–27.

[15] O. Boiman. "In defense of nearest neighbor based image classification", In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2008, Anchorage, Alaska, USA. p. 1-8.

[16] L. A. Zadeh. "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," AI magazine, 1986, vol. 2, no. 7, p. 85-90.

[17] J. Dezert, P. Wang, A. Tchamova, "On The Validity of Dempster-Shafer Theory", in Proc. of Fusion 2012 Int. Conference, Singapore, July 2012.

[18] A. Tchamova, J. Dezert, "On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory", in Proc. of the 6th IEEE International Conference on Intelligent Systems, Sofia, Bulgaria, Sept. 2012.

[19] P. Smets.. "The transferable belief model," Artificial Intelligence, 1994, vol. 66, p. 191-234.

[20] J.-M. Tacnet, J. Dezert. "Cautious OWA and evidential reasoning for decision making under uncertainty," in Proc. of 14 International Conference on Information Fusion, July 5-8, 2011, Chicago, USA, p. 2074–2081.

[21] D. Q. Han, J. Dezert, J.-M. Tacnet, C. Z. Han. "A fuzzy cautious OWA approach with evidential reasoning," in Proc. of 15 International Conference on Information Fusion, July 9-12, 2012, Singapore, p. 278-285.

[22] L.I. Kuncheva, C.J. Whitaker, "Measures of diversity in classifier ensembles", Machine Learning, Pages 181–207, Volume 51, Issue 2, May 2003.