A new belief-based K-nearest neighbor classification method

Zhun-ga Liu a, Quan Pan a,*, Jean Dezert b

a School of Automation, Northwestern Polytechnical University, Xi’an, China
b ONERA - The French Aerospace Lab, F-91761 Palaiseau, France

ARTICLE INFO

Article history:
Received 1 November 2011
Revised in revised form 2 August 2012
Accepted 1 October 2012
Available online 11 October 2012

Keywords:
K-nearest neighbor
Data classification
Belief functions
DST
Credal classification

ABSTRACT

The K-nearest neighbor (K-NN) classification method originally developed in the probabilistic framework has serious difficulties to classify correctly the close data points (objects) originating from different classes. To cope with such difficult problem and make the classification result more robust to misclassification errors, we propose a new belief-based K-nearest neighbor (BK-NN) method that allows each object to belong both to the specific classes and to the sets of classes with different masses of belief. BK-NN is able to provide a hyper-credal classification on the specific classes, the rejection classes and the meta-classes as well. Thus, the objects hard to classify correctly are automatically committed to a meta-class or to a rejection class, which can reduce the misclassification errors. The basic belief assignment (bba) of each object is defined from the distance between the object and its neighbors and from the acceptance and rejection thresholds. The bba’s are combined using a new combination method specially developed for the BK-NN. Several experiments based on simulated and real data sets have been carried out to evaluate the performances of the BK-NN method with respect to several classical K-NN approaches.

Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.

1. Introduction

The K-nearest neighbor (K-NN) method [13,14,27,1] is a well known classification algorithm used in pattern recognition. In the original voting K-NN, the object is assigned to the majority class according to its K nearest neighbors (KNNs), and the distances between the object and its neighbors are ignored. In [11], a weighted version of K-NN (WK-NN) taking into account the distance between the object and its KNNs has been proposed by Dudani, and it can provide better performance than voting K-NN method in the data classification. In the classification of close data sets, some similar data points originating from different classes are difficult to classify correctly into a specific class and we call them as imprecise data. In such case, the classical K-NN methods cannot well deal with ambiguity and imprecise information because of the limitation of the probabilistic framework, more precisely its inability to make a clear distinction between the full lack of knowledge (circumvented by the insufficient reason argument) and the full knowledge of the equiprobable cases.

Belief functions theory [26,28–30] also called evidence theory (or Dempster–Shafer Theory (DST)) makes a clear distinction between the different aspects of uncertainty (mainly the randomness and the non-specificity) through the rigorous formalism of belief functions. DST has been already applied in many fields, including parameter estimation [5], classification [6,7,9,17,18,21,25,31], clustering [2,8,19,22–24] and decision-making support [20]. Credal partition (classification) as an extension of the hard classification and the probabilistic (fuzzy) classification has been introduced for the data clustering under the belief functions framework in [8,22]. It allows the objects to belong to not only the specific classes but also the set of classes (i.e. meta-class) with different masses of belief. The imprecise data that hard to classify correctly in specific classes, are more reasonably committed to the meta-class defined by the disjunction of some specific classes they are close to, which can reduce (avoid) the misclassification errors.

The K-NN has been extended in the belief functions framework by Denœux [6] to better model the uncertain information. Denœux’s method is referred as the evidential K-NN method (or EK-NN for short) in this paper. In EK-NN, a simple construction of the basic belief assignments (bba’s) of the object associated to its KNNs was proposed, and only two focal elements (i.e. the specific class and total ignorant class) are involved in each bba. The combination results of these bba’s by Dempster’s rule (also called DS rule for short) are used to classify the objects. In EK-NN, it works only with one extra focal element (representing the total ignorant class) with respect to what is done within the classical K-NN approach. The meta-classes (i.e. the partial ignorant classes), which can well represent the imprecise data, are not considered...
as possible classification results in EK-NN. Moreover, it is worth to note that even if one object is far from its KNNs, after the combination of the K bba’s the largest mass of belief can be committed to a specific class rather to the total ignorance due to the particular structure of bba’s in EK-NN, which is not a good expected behavior.

A new data classification method called belief-based K-nearest neighbor (BK-NN) is proposed here to overcome the limitations of EK-NN, and the meta-class is taken into account to characterize the imprecise data. BK-NN can reduce misclassification errors by making the classification result partly imprecise and in the meantime robust thanks to the introduction of meta-classes. In making the classification result partly imprecise and in the imprecise data. BK-NN can reduce misclassification errors by EK-NN, and the meta-class is taken into account to characterize the neighbor (BK-NN) is proposed here to overcome the limitations of K-NN. The specific class refers to the data (object) very close to the other classes in the frame of threshold) enough to the other classes. If one object is in a combination of the

For example, in the handwritten digits (e.g. USPS data set [161]) classification problem, some badly written digits are very difficult to recognize, and they should be prudently committed to the associated meta-classes or simply be regarded as an outlier in the extreme case in order to reduce the misclassification errors. Let us consider a simple 2-class of handwritten digits 1 and 4 in USPS data set. The two badly written digits of 4 as Fig. 1(a) and (b) are similar but not enough to be associated easily with a common handwritten digit 4. The samples 1 and 2 look also quite similar to the written digit 1. In such a difficult case, it is better to commit the samples 1 and 2 to the meta-class {1,4} to avoid a wrong precise classification. The samples 3 and 4 of handwritten digit 4 shown in Fig. 1(c) and (d) are almost impossible to classify visually as corresponding to the digit 4, and they are neither similar to the digit 1. So these two samples can be considered as an outlier. Such badly handwritten digits belonging to the meta-class or outlier class should be classified manually by human operators or by other methods with additional information.

In BK-NN, the bba’s of one object is constructed according to the distances between the object and its KNNs and two (acceptance and rejection) thresholds. Three focal elements including specific class, the rejection of this specific class (called rejection class), and the ignorant class are involved in the construction of the bba’s. The potential unknown (outlier) class is considered being included in the rejection classes. To deal efficiently with the imprecise (ambiguous) data hard to classify, the meta-classes are conditionally selected in the fusion process of the bba’s according to the number of their KNNs in each class. A new combination rule inspired by DS rule [26] and Dubois–Prade (DP) rule [10] is proposed to combine these bba’s, and some partial conflicting beliefs are preserved and committed to the selected meta-classes. Three kinds of classes, which include specific classes, rejection classes and meta-classes, will be involved in the fusion results and this constitutes the hyper-credal classification. Thus, BK-NN works with a more enlarged classification spectrum than EK-NN and K-NN. The specific class refers to the data (object) very close to the training samples in this specific class as in K-NN. The rejection class defined by the complement of several classes refers to data that is quite far (beyond the rejection threshold) from these several classes, but not close (beyond the acceptance threshold) enough to the other classes. If one object is in a rejection class, we can just confirm that it is not in the several specific classes involved in the rejection class, but it does not indicates that it must belong to the other classes in the frame of discernment, since it still could be in the potential outlier class. The meta-class defined by the disjunction of several specific classes refers to the data that is simultaneously close to all the involved specific classes, which are not very distinguishable for the data in this meta-class. If some objects are considered in the meta-classes or in the rejection classes, it means that the available information is not sufficient enough for making a precise (specific) classification of such objects. Actually, the output of BK-NN can also viewed as an interesting tool to justify the use of complementary information sources if very precise classification results are sought whenever BK-NN fails to provide them. In the next section, the background of belief functions theory is briefly introduced. In Section 3, EK-NN method is briefly recalled and analyzed. The details of BK-NN are presented in Section 4. Several experiments are given in Section 5 to show how BK-NN performs with respect to EK-NN. Concluding remarks are given in the last section of this paper.

2. Basics of belief functions theory

Belief functions theory is also called Dempster–Shafer theory (DST), or evidence theory [26,28–30]. Let us consider Θ being a finite discrete set of mutually exclusive and exhaustive hypotheses. Θ is called the frame of discernment of the problem under consideration. The power-set of Θ, 2Θ, contains all the subsets of Θ. The credal partition [8,22] for data clustering over the frame Θ corresponds to the power-set 2Θ.

For example, if Θ = {θ1,θ2,θ3}, then 2Θ = {∅,θ1,θ2,θ3,θ1∪θ2, θ1∪θ3,θ2∪θ3,Θ}. A basic belief assignment (bba) is a function m(·) from 2Θ to [0,1] satisfying

\[ \sum_{A \subseteq 2^\Theta} m(A) = 1 \]
\[ m(\emptyset) = 0 \]  (1)

The quantity m(·) can be interpreted as a measure of the belief that is committed exactly to A, and not to any of its subsets. The subsets A of Θ such that m(A) > 0 are called the focal elements of m(·), and the set of all its focal elements is called the core of m(·).

Several distinct bodies of evidence characterized by different bba’s can be combined using DS rule [26]. Mathematically, the DS rule of combination of two bba’s m1(·) and m2(·) defined on 2Θ is defined2 by mDS(0) = 0 and for A ≠ 0, B,C ∈ 2Θ by

\[ m_{DS}(A) = \frac{\sum_{B \subseteq C} A m_1(B)m_2(C)}{\sum_{B \subseteq C} A m_1(B)m_2(C)} = \frac{\sum_{B \subseteq C} A m_1(B)m_2(C)}{\sum_{A \subseteq C} m_1(A)m_2(C)} \]  (2)

In DS rule, the total conflicting belief mass \( \sum_{B \subseteq C} A m_1(B)m_2(C) \) is redistributed back to all the focal elements through the normalization.

To palliate some drawbacks of DS rule, Dubois and Prade have proposed their own rule in [10], denoted DP rule in this paper. The basic idea behind DP rule consists to transfer each partial

---

2 Which means that most of the mass of belief was committed to the ignorance in each bba’s to combine.

3 This rule is mathematically defined only if the denominator is strictly positive, i.e. the sources are not totally conflicting.
conflicting mass on the disjunction of the elements involved in the partial conflict. Mathematically, DP rule is defined by

\[ m_{DP}(A) = \sum_{B \subseteq A} m(B)m_2(C) + \sum_{k \subseteq A} m_k(B)m_2(C) \]  

(3)

The belief function \( Bel(A) \) and the plausibility function \( Pl(A) \) are usually interpreted as lower and upper probabilities of \( A \). They are mathematically defined by

\[ Bel(A) = \sum_{A \subseteq B \subseteq A} m(B) \]  

(4)

\[ Pl(A) = \sum_{A \subseteq B \subseteq 2^n} m(B) \]  

(5)

Pignistic probability transformation \( BetP(\cdot) \) (betting probability) \[29,30] \] is commonly chosen to transform any bba \( m(\cdot) \) into a probability measure for the decision-making support in the framework of belief functions. Mathematically, \( BetP(A) \) is defined for \( A \in 2^n \cup \{0\} \) by

\[ BetP(A) = \sum_{A \subseteq B \subseteq 2^n} \frac{|A \cap B|}{|B|} m(B) \]  

(6)

where \( |X| \) is the cardinality of the element \( X \).

3. Brief recall and comments on EK-NN

The original voting K-NN \[13\] is a well-known non-parametric classification method used in the Pattern Recognition field. In this method, an object to classify is assigned to the class represented by a majority of its KNNs in the training set. It has been pointed by Cover and Hart in \[3\] that the error rate of the \( K \)-NN rule tends to infinity as \( K/N \to 0 \). However, the voting \( K \)-NN rule is not guaranteed to be the optimal way of using the information contained in the neighborhood of unclassified patterns in the finite sample case. That is why many attempts have been done to improve K-NN rule, like distance-weighted \( K \)-NN rule \[11\], etc. An evidential K-nearest neighbor (EK-NN) method was proposed by Denœux in \[6\] based on DST, and it can provide a global treatment of some important issues: the consideration of the distances information, ambiguity and distance rejection, and the consideration of uncertainty.

Let us consider a group of objects \( X = \{x_1, x_2, \ldots, x_n\} \) to be classified into several classes, and the frame of the classes is \( \Omega = \{\omega_1, \ldots, \omega_l\} \). In EK-NN \[6\], the bba's of the object \( x_i \) associated to the training data \( x_i \) with labeled class \( \omega_j \in \Omega \) is defined as

\[ m_1^i(\omega_j) = \alpha e^{-\gamma d_i^{\beta}} \]  

(7)

\[ m_2^i(\omega_j) = 1 - e^{-\gamma d_i^{\beta}} \]  

(8)

where \( \alpha, \gamma, \beta \) and \( \beta \) are tuning parameters, and \( d \) is the distance between the objects and the training data. The justification of the choice of these bba's is given by Denœux in \[6\].

As shown through formulas \( (7)-(8) \), in EK-NN only the two focal elements \( \omega_j \) and \( \Omega \) are involved for a given bba \( m^i(\cdot) \), \( i = 1, \ldots, n \). The classification result of each object \( x_i \) is obtained from the fusion of the K bba's associated with the K nearest neighbors using DS rule \[26\]. Because of the particular structure of bba's, the EK-NN based on DS rule can only produce the specific classes and the total ignorance class \( \Omega \) as focal elements. Therefore, no meta-classes (partial ignorant classes) like \( \omega_j \cup \cdots \cup \omega_k \equiv [\omega_1, \ldots, \omega_l] \) can be a focal element, which is not convenient for dealing efficiently with imprecise (non specific) information.

The particular structure of bba's chosen in EK-NN generates also a serious problem because when \( K \) is big enough and if the \( K \)-nearest neighbors of the object have the same labeled class \( \omega_i \) but are quite far from it, the object will be committed to the class \( \omega_i \) rather than to the ignorance \( \Omega \). Such EK-NN behavior is unreasonable as we will show in the following example.

**Example 1.** Let us assume that one object \( x_i \) is very far from all the training data, and its \( K = p \) nearest neighbors belong to the same class \( \omega_i \), and they are located at the same distance from \( x_i \), so that we must consider the following \( p \) bba's, with \( \delta \in [0,1] \):

\[ \begin{align*}
    m_1(\omega_j) &= \delta \\
    m_1(\Omega) &= 1 - \delta
\end{align*} \]  

(9)

where \( i = 1, \ldots, p \).

Then, the fusion result obtained by DS rule is

\[ m(\omega_1) = 1 - (1 - \delta)^p, \quad m(\Omega) = (1 - \delta)^p \]  

(10)

One can see that \( m(\omega_1) \) increases when \( p \) increases. So even if the value \( \delta > 0 \) is very small, which means that \( x_i \) is far from its \( K \)-nearest neighbors, then the mass of belief committed to \( \omega_1 \) can be very large when \( p \) is big enough. It can be easily verified that when \( p > 10 \), then \( m(\omega_1) > 0.99 \). For example, if \( \delta = 0.2 \), and \( p > 3 \), then \( m(\omega_1) > 0.5 \). Such behavior of EK-NN approach due to the particular structure of bba's to combine and the mechanism of the DS rule is not very reasonable, nor appropriate in fact. In our opinions, the object \( x_i \) should be classified as outlier when it is far from its \( K \)-nearest neighbors, no matter how many nearest neighbors are given.

4. Belief-based K-nearest neighbor method (BK-NN)

A new data classification method called belief-based K-nearest neighbor method (BK-NN) is proposed to overcome the limitations of EK-NN. In BK-NN, the \( K \) bba's associated with one object and its KNN's are computed from the distance between the object and its neighbors, and from the acceptance and rejection thresholds in each class. Thus, each bba contains three focal elements including the specific class \( (\text{e.g.} \omega_i) \), the rejection class \( (\text{e.g.} \omega_j) \) and the ignorant class \( \Omega \). The potential unknown (outlier) class is taken into account in the rejection class. This new structure of bba's is able to better reveal the degree of one object belonging (or not) to a class label of its close neighbor.

In the fusion of these bba's, if several specific classes appear undistinguishable for one object according to the number of its KNN's in each class, the meta-class (the disjunction of these specific classes) will be taken into account in the fusion process to avoid false precise classification. The associated partial conflicting beliefs will be kept and committed to the selected available meta-classes by a new fusion rule inspired from DP rule. Hence, BK-NN allows to classify an object not only to a specific classes, but possibly also to the sets of classes (i.e. meta-classes and rejection classes) with the different masses of belief. This enlarged classification, called hyper-credal classification, provides a deeper insight into the data, and allow to reduce the error rate in the classification results.

4.1. The construction of bba's in BK-NN

Let us consider the frame of the classes as \( \Omega = \{\omega_0, \omega_1, \ldots, \omega_l\} \) including one potential unknown (outlier) class \( \omega_0 \). The \( K \) nearest neighbors of the data point \( (\text{object}) \) \( x_i \) are found at first as with the K-NN method. The acceptance and rejection thresholds with respect to the distance between the object and its KNN's are introduced for the construction of bba's as illustrated in Fig. 2.
We assume that one of the K-nearest neighbors $x_i$ of the object $x_t$ to classify is labeled by class $w$. In BK-NN, the determination of the bba's associated with $x_t$ and $x_i$ is based on the given acceptance threshold $d_{ws}^m$ and the given rejection threshold $d_{ta}^m$ with $d_{ta}^m < d_{ws}^m$ given for each class. If the distance $d_{ij}$ between $x_i$ and $x_t$ is smaller than $d_{ta}^m$, it indicates that $x_i$ lies in the acceptance zone of $w$, with respect to $x_t$, and $x_i$ will be considered most likely to belong to class $w$. If $d_{ij}$ is larger than $d_{ws}^m$, it means that $x_i$ is beyond the rejection zone. So $x_i$ does not very likely belong to the class $w$, and it has more chance to belong to the complementary set (rejection) of $w$, denoted $\overline{w}$, rather than $w$. If $d_{ta}^m \leq d_{ij} \leq d_{ws}^m$, we consider that $x_i$ can be reasonably committed to uncertainty (i.e. the full ignorance $w \cup \overline{w}$) because there is no strong support in favor of $w$ w.r.t. $\overline{w}$, and conversely.

In BK-NN, the bba's are defined as a function of the distance between the object and the training data, and three focal elements are involved in the construction of the bba's. This makes a big difference w.r.t. what was proposed in EK-NN approach as it will be shown in the sequel. So, for any object $x_t$, its bba's associated with one of its K-nearest neighbors $x_i$ will be obtained by the fusion of the two elementary bba's $m_1(\cdot|d_{ij})$ and $m_2(\cdot|d_{ij})$ listed in the following table:

| Focal element | $m_1(\cdot|d_{ij})$ | $m_2(\cdot|d_{ij})$ |
|---------------|---------------------|---------------------|
| $w$           | $f_1(d_{ij},d_{ws}^m)$ | $0$ |
| $\overline{w}$ | $0$ | $f_2(d_{ij},d_{ta}^m)$ |
| $\Omega$     | $1-f_1(d_{ij},d_{ws}^m)$ | $1-f_2(d_{ij},d_{ta}^m)$ |

$d_{ij} \equiv d(x_i, x_t)$ is the distance between the object $x_i$ and its closest neighbor $x_t$. The functions $f_1(\cdot)$ and $f_2(\cdot)$ characterizing the good (expected) behavior for an efficient bba to a class, its complement, or to ignorance should satisfy the following properties:

1. $f_1(\cdot)$ should be monotone decreasing, and $f_1(\cdot) \in [0,1]$;
2. If $d_{ij} \leq d_{ws}^m$, then $f_1(d_{ij},d_{ws}^m) \geq 0.5$;
3. $f_2(\cdot)$ should be monotone increasing, and $f_2(\cdot) \in [0,1]$;
4. If $d_{ij} \geq d_{ta}^m$, then $f_2(d_{ij},d_{ta}^m) \geq 0.5$.

It has been shown in [4] that the following two sigmoid functions satisfy the aforementioned requested properties and thus they can be used to construct the bba's $m_1(\cdot|d_{ij})$ and $m_2(\cdot|d_{ij})$:

$$f_1(d_{ij},d_{ws}^m) = \frac{1}{1+e^{-\lambda_1(d_{ij}-d_{ws}^m)}}$$

$$f_2(d_{ij},d_{ta}^m) = \frac{1}{1+e^{-\lambda_2(d_{ij}-d_{ta}^m)}}$$

$\lambda_1/4$ is the slope of the tangent at the inflection point. $d_{ws}^m$ and $d_{ta}^m$ are the abscissa of the inflexion point of the sigmoid. In the determination of these tuning parameters, the average distance between each training data in $w$ and its KNNS can be calculated first, and the mean value $d_{ws}^m$ of all the average distances for the training data in $w$ can be used to determine $d_{ws}^m$, $d_{ta}^m$ and $\lambda$. The bigger $d_{ws}^m$ usually leads to the smaller $\lambda$, and one suggests to take $\lambda_1 = \frac{1}{d_{ws}^m}$

The values $d_{ws}^m$ and $d_{ta}^m$ can be chosen as

$$[d_{ws}^m, d_{ta}^m] = [\gamma_{1,0}, \gamma_{1,1}] \times d^{m}$$

We recommend to take $\gamma_{1,0} \in [1.3]$ and $\gamma_{1,1} \in [2.1, 2.3]$, accordingly to our applications. The bigger $d_{ws}^m$ will cause more objects committed into the meta-classes, and the smaller $d_{ta}^m$ will make more objects considered as outliers. $[d_{ws}^m, d_{ta}^m]$ is the imprecise zone of the class $w$. The bigger gap between $d_{ws}^m$ and $d_{ta}^m$ will lead to more objects in the rejection classes. In fact, the value of $d_{ws}^m$ and $d_{ta}^m$ can be optimized using the training data, and it will be presented in sequel.

It can be verified that the bba $m_1^w(\cdot|d_{ij})$ associated with the point $x_i$ and class $w$, satisfying the global expected behavior is then obtained by the fusion of the two simple bba's $m_1(\cdot|d_{ij})$ and $m_2(\cdot|d_{ij})$. Because of the structure of these two simple bba's $m_1(\cdot|d_{ij})$ and $m_2(\cdot|d_{ij})$, the fusion obtained with DS rule [26], PRS rule [28], and even with DSm hybrid (DSmH) rule [28] provide globally similar results and therefore the choice of the fusion rule here does not really matter to build $m^w(\cdot|d_{ij})$. For simplicity, DS rule (2) is used here to combine $m_1(\cdot|d_{ij})$ and $m_2(\cdot|d_{ij})$, for getting $m^w(\cdot|d_{ij})$ since for historical reasons, it is the most common combination rule used so far in the belief functions framework.

Based on this approach, K bba's corresponding to KNNs of $x_i$ can be constructed. Moreover, if there is no point in the KNNS of $x_i$ belonging to class $w_p$, we consider that $x_i$ is impossible to belong to class $w_p$, and in that case the additional particular bba's should be added as

$$m^w(v_{wp}) = 1$$

This particular bba's is added to eliminate $w_p$ from the possible classification results. For example, let us consider $\Omega = \{w_0, w_1, w_2\}$. If the KNNS of one object are all in class $w_1$, but this object is quite far from the training data. Then the focal element $\overline{w_1}$ will take the most mass of belief in the construction of the K bba's. If the particular bba's $m^w(v_{wp}) = 1$ is not included, $\overline{w_1} = \{w_0, w_2\}$ will still take the biggest mass of belief in the combination results of the K bba's. However, we consider that this object should belong to the outlier class $\overline{w_1} \cup \overline{w_2} = w_0$. Thus, the particular bba's is necessary to be fused with the combination results of the K bba's in order to remove the impossible class $w_1$ as $\overline{w_1} \cup w_2 = \overline{w_0}$.

It is worth to note that if $x_i \in \overline{w_2}$ means that $x_i$ is not in the class $w_2$, but it does not mean that $x_i$ belongs to the class $w_2$, since $x_i$ can still be not close enough to $w_2$ (beyond the acceptance threshold

---

4 Of course, some other proper functions (like truncated trapezoidal functions by example) can also be selected (designed) by the users according to their actual applications.

5 The choice of the combination rule is not the main point of this paper. We prefer to focus on the general presentation of the method.
of \( w_2 \), and therefore \( w_2 \) could be in fact an outlier, that is why we consider \( \overrightarrow{w_1} = \{ w_2, w_0 \} \neq w_2 \) in BK-NN approach.

4.2. The combination of bba's in BK-NN

The fusion results of the K bba's (with the additional particular bba's if available) for one object can be used to classify this object. The fusion process of the K bba's consists of two steps: (1) subcombination of the bba's associated with the same class; (2) global fusion of the sub-combination results about different classes.

In the sub-combination of the bba's associated with the close neighbors labeled by the same class, the conflicting beliefs (e.g. \( m_1(w_1) \cdot m_2(w_2) \)) produced by the conjunction of the bba's do not play a particular role, and they must be distributed to the other focal elements to preserve the normalization of the combined bba. So DS rule (2) can be used here for the combination of the bba's in this first step.

The sub-combination results \( m^{\text{conf}}(\cdot) \), for \( s = 1, \ldots, h \) (h being the number of the classes in the frame of discernment \( \Omega \)) issued from the first step will then enter in the global fusion process. If DS rule is used for the global fusion of \( m^{\text{conf}}(\cdot) \), \( s = 1, \ldots, h \), all the conflicting beliefs about the conjunction of different classes (e.g. \( m^{\text{conf}}(w_1) \cdot m^{\text{conf}}(w_2) \)) will be redistributed. Consequently, the imprecise objects having KNNS with different classes label and having approximately the same number of the KNNS in each class will be mostly wrongly classified. To illustrate this problem, let us consider one object having 11-nearest neighbors with six neighbors in class \( w_1 \) and five neighbors in class \( w_2 \). Such object is usually considered very difficult to classify into one of the particular (precise) classes \( w_1 \) or \( w_2 \). Intuitively, it is more reasonable and prudent to commit such object to the meta-classes \( w_1 \cup w_2 \) representing the uncertainty one has, which can reduce the risk of false classification. The partial conflicting masses (e.g. \( m^{\text{conf}}(w_1) \cdot m^{\text{conf}}(w_2) \)) capture the degree of the objects belonging to the corresponding meta-classes (e.g. \( w_1 \cup w_2 \)). If all the conflicting masses are simply kept and committed to the associated meta-classes as done by DP rule, it will cause very large number of meta-classes especially when \( K \) is big, and too many objects will be classified into the meta-classes, which is not an efficient data classification solution. For example, let us choose \( K = 11 \). One object has 10 nearest neighbors in class \( w_1 \) and one neighbor in class \( w_2 \). If the 11 neighbors are all close to the object according to the given acceptance threshold, this means that the class \( w_1 \) will get highest belief from the 10 bba's satisfying BK-NN bba modeling, whereas \( w_2 \) will get highest belief from the other bba. Then the meta-class \( w_1 \cup w_2 \) will get the most belief after the combination of the 11 bba's by DP rule because of the existing partial conflict \( m^{\text{conf}}(w_1) \cdot m^{\text{conf}}(w_2) \). However, this object is more likely in \( w_1 \) than in \( w_2 \), since there are 10 neighbors in \( w_1 \), and only one neighbor belonging to \( w_2 \). So it seems not very reasonable to commit this object into the meta-class \( w_1 \cup w_2 \) as obtained when applying DP rule. That is why we propose a new global fusion rule which is a hybridization of DS and DP rules.

In order to reach a good compromise between the imprecision rate and the error rate for classification, we argue that the partial conflicting beliefs corresponding to the meta-class should be conditionally preserved according to current context.

The KNNS of the imprecise objects usually have different class labels, and the number of the KNNS in different classes is similar. So the meta-classes can be automatically selected according to the number of KNNS in different classes for one object.

Let us consider the KNNS of one object are associated with \( w_1, \ldots, w_h \), and the number of the KNNS in each class is \( k_1, \ldots, k_h \). The biggest value \( k_{\text{max}} = \max(k_1, \ldots, k_h) \) is found at first. Then if \( k_{\text{max}} - k_i \leq t, i = 1, \ldots, g \) (t being a given threshold), the object will be very likely in the union of classes (also called meta-class) \( \psi_{\text{max}} = \{ w_l | k_{\text{max}} - k_i \leq t \} \), and \( w_i = 1, \ldots, g \) are potentially indistinguishable for the object since the number of the KNNS in \( w_i = 1, \ldots, g \) is similar. However, the number of KNNS in each class is not sufficient to determine the class of the object, and the object can possibly be in a class which is not included in \( \psi_{\text{max}} \) in rare cases. To deal with this problem, all the meta-classes having a cardinality\(^6\) less than or equal to \( \psi_{\text{max}} \) will be used. By doing this, all the classes in \( \Omega \) can be fairly treated. The set of the selected meta-classes is denoted by \( \Psi \). For example, let us consider \( \Omega = \{ w_0, w_1, w_2, w_3 \} \) with the number of KNNS in each corresponding class is \( k_1, k_2, k_3 \). If \( k_{\text{max}} = k_2 \) and \( \psi_{\text{max}} = \{ w_2, w_3 \} \), then all the meta-classes whose cardinality is not bigger than \( \{ w_2, w_3 \} \) will be selected, and the selected meta-classes are given by \( \Psi = \{ w_1 \cup w_2, w_2 \cup w_3 \} \). If the number of KNNS in \( w_{\text{max}} \) is much bigger than that in the other classes as \( k_{\text{max}} - k_i > t, i \neq \text{max} \), none meta-class needs to be preserved in order to avoid the high imprecision of the solution.

The tuning threshold \( t \) should become bigger with the increasing of the chosen number of nearest neighbors. So it can be simply determined by

\[
t = \max(1, \eta \times K)
\]

where we recommend to take \( \eta \in [0.1, 0.3] \) and \( t \geq 1 \). The bigger \( t \) can produce the fewer misclassification errors, but it usually leads to the larger meta-classes. \( t \) should be determined according to the expected compromise we want between the imprecision and misclassification of the results.

The global fusion rule inspired from DS rule (2) and DP rule (3) is mathematically achieved by the formulas (17) and (18). The bba's of the sub-combination results about the different classes can be fused sequentially by

\[
m_{\text{fused}}(A) = \left\{ \begin{array}{l}
\sum_{A_1 \in A} m_{\text{unfused}}(A_1) m_{\text{conf}}(B_2)
\end{array} \right.
\]

\[
m_{\text{unfused}}(A)
\]

where \( m_{\text{unfused}}(\cdot) \) is the unnormalized combination results of \( m_1(\cdot), \ldots, m_h(\cdot) \), and \( m_1(\cdot) = m^{\text{conf}}(\cdot) \). It is worth to note that this combination rule is close to DP rule (3), but the summation of the combined bba is not one (i.e. here one can have \( \sum_{A_1 \in A} m_{\text{unfused}}(A_1) \leq 1 \)) if some partial conflicting beliefs are not preserved. In DP rule, the focal element \( A \) can be any subset of \( \Omega \), whereas in our rule a focal element \( A \) can be only a specific class, the ignorant class or just a selected meta-class (but not all possible meta-classes). This unnormalized combination rule is associative here because of the particular structure of bba's \( m^{\text{conf}}(\cdot) \) satisfying the BK-NN modeling, and the combination order of the bba's has no effect on the results.

The unnormalized fusion results obtained by (17) will then be normalized once all the bba's have been combined. The conflicting masses that are not committed to the meta-classes will be redistributed to the other focal elements. There are many ways to redistribute the conflicting masses. They can be redistributed to all the focal elements as done by the DS rule, or they can be redistributed to particular focal elements as done by the Proportional Conflict Redistribution Rule no. 5 (PCR5) rule [28] (Vol. 2), etc. However, the redistribution method does not matter too much here on the fusion results, since the conflicting masses to be redistributed are usually very small, and the majority of the conflicting masses are generally committed to the selected meta-classes. For reducing the computation burden, we use the simple

\[
\text{The cardinality of a set is the number of the elements it contains. For example, } |w_i \cup w_j| = 2.
\]
normalization of the fusion results given by

\[
m(\Lambda) = \frac{m_{\Lambda}(\Lambda)}{\sum m_{\Lambda}(B_i)} \tag{18}
\]

If none meta-class is selected, the global fusion rule will reduce to DS rule. Whereas, if all the meta-classes are preserved, the global fusion rule behaves like DP rule, and all the partial conflicting beliefs are kept for these meta-classes. This global fusion rule can be considered as a compromise between DS rule and DP rule, since there are just part of the meta-classes selected to be available depending on the actual condition.

The global fusion results can be used for the classification making support, and the belief function \( Be_k(\cdot) \), the plausibility function \( Pl_k(\cdot) \) and pignistic probability \( BetP_k(\cdot) \) introduced in Section 2 can also be used for decision making if necessary.

Three tuning parameters \( \gamma_k \), \( \gamma_t \), and \( \eta \) are necessary in the BK-NN algorithm. The parameters \( \gamma_k \), \( \gamma_t \) can be obtained by a cross-validation using training data. The specific classification of each object is determined according to the classical pignistic probability \( BetP(\cdot) \) which allows to transform any bba into a probability measure. One can keep a validation set, and make sure the specific classification of the objects as correct as possible by tuning these two parameters for the given value of \( K \) nearest neighbors. The optimized value of the parameters corresponding to the lowest error rate can be found over a grid search. Based on our experiences and tests, one recommends the search scope as \( \gamma_k \in [1,3] \) and \( \gamma_t \in [2/3,3/2] \). The parameter \( \eta \) entering in the computation of the threshold \( t \) given in (16) allows to control the number of the objects in the meta-classes. This parameter \( \eta \) has however only a very little influence on the hard classification results of the objects. So \( \eta \) should be tuned depending on the imprecision degree (i.e. the number of the objects in the meta-classes) one wants to accept.

The general principle of BK-NN for data classification is summarized in the flowchart of the Fig. 3.

**Example 2.** Let us consider the following frame of classes \( \Omega = \{w_0, w_1, w_2, w_3\} \), and an object \( x \) with 5-nearest neighbors including three neighbors labeled by classes \( w_1 \) and two neighbors by \( w_0 \) with the corresponding bba’s satisfying BK-NN bba modeling:

\[
m_{\Omega}(w_1) = 0.8, \quad m_{\Omega}(\Omega) = 0.1, \quad m_{\Omega}(w_0) = 0.1
\]

\[
m_{\Omega}(w_1) = 0.6, \quad m_{\Omega}(\Omega) = 0.3, \quad m_{\Omega}(w_0) = 0.1
\]

\[
m_{\Omega}(w_1) = 0.4, \quad m_{\Omega}(\Omega) = 0.4, \quad m_{\Omega}(w_0) = 0.2
\]

\[
m_{\Omega}(w_2) = 0.5, \quad m_{\Omega}(\Omega) = 0.2, \quad m_{\Omega}(w_0) = 0.3
\]

\[
m_{\Omega}(w_0) = 0.6, \quad m_{\Omega}(\Omega) = 0.1, \quad m_{\Omega}(w_0) = 0.3
\]

It is worth to note that we assume \( \omega = w_0 \cup w_1 \cup w_2, \omega = w_0 \cup w_1 \cup w_3 \) and \( \omega = w_0 \cup w_1 \cup w_2 \) in the calculation of the mass of belief committed to the focal elements \( A, A \subseteq \Omega \).

We can see that none of the 5-nearest neighbors is in class \( w_0 \), which indicates that \( x \) is impossible to belong to \( w_0 \). So one particular bba’s should be additionally introduced as

\[
m_{\Omega}(\Omega) = 1
\]

\[
m_{\Omega}(w_1), m_{\Omega}(w_2), m_{\Omega}(\Omega) \quad \mbox{associated with } \quad w_1, \quad \mbox{and } \quad m_{\Omega}(w_0), m_{\Omega}(w_0) \quad \mbox{associated with } \quad w_2,
\]

are respectively combined using DS rule given by Eq. (2). The combination results are

\[
m_{\Omega}(w_1) = 0.8867, \quad m_{\Omega}(\Omega) = 0.1085, \quad m_{\Omega}(w_0) = 0.0048
\]

\[
m_{\Omega}(w_2) = 0.7590, \quad m_{\Omega}(\Omega) = 0.1325, \quad m_{\Omega}(w_0) = 0.1085
\]

We get \( k_{\max} = k_1 = k_2 = 3 \). If we select \( t = 2 \), then \( |k_1 - k_2| = 3 - 2 < 2 \) and \( |k_1 - k_2| = |3 - 0| = 3 \). So \( \psi_{\max} = \{w_1, w_2\} \), and the selected meta-classes are \( \Psi = \{w_1 \cup w_2, w_3 \cup w_1, w_2 \cup w_3\} \). The global fusion results of \( m_{\Omega}(\cdot), m_{\Omega}(\cdot) \) and \( m_{\Omega}(\cdot) \) using formulas (17) are given by

\[
m_{\Omega}(w_1) = \left[m_{\Omega}(w_1) m_{\Omega}(\Omega) + m_{\Omega}(w_1) m_{\Omega}(w_0)\right] \cdot m_{\Omega}(w_1) = 0.2137
\]

\[
m_{\Omega}(w_2) = \left[m_{\Omega}(w_2) m_{\Omega}(\Omega) + m_{\Omega}(w_2) m_{\Omega}(w_0)\right] \cdot m_{\Omega}(w_2) = 0.0860
\]

\[
m_{\Omega}(w_1 \cup w_2) = m_{\Omega}(w_1) m_{\Omega}(w_2) m_{\Omega}(w_1 m_{\Omega}(w_2)) = 0.6730
\]

\[
m_{\Omega}(w_1 \cup w_2 \cup w_3) = \left[m_{\Omega}(w_1) m_{\Omega}(w_2) m_{\Omega}(\Omega)\right] \cdot m_{\Omega}(w_1 \cup w_2 \cup w_3) = 0.0005
\]

\[
m_{\Omega}(w_1 \cup w_2 \cup w_3) = \left[m_{\Omega}(w_1) m_{\Omega}(w_2) m_{\Omega}(\Omega)\right] \cdot m_{\Omega}(w_1 \cup w_2 \cup w_3) = 0.0118
\]

\[
m_{\Omega}(w_1 \cup w_2 \cup w_3) = \left[m_{\Omega}(w_1) m_{\Omega}(w_2) m_{\Omega}(\Omega)\right] \cdot m_{\Omega}(w_1 \cup w_2 \cup w_3) = 0.0006
\]

\[
m_{\Omega}(w_1 \cup w_2 \cup w_3) = \left[m_{\Omega}(w_1) m_{\Omega}(w_2) m_{\Omega}(\Omega)\right] \cdot m_{\Omega}(w_1 \cup w_2 \cup w_3) = 0.0144
\]

The combination results have been normalized, since there is no conflicting beliefs to be redistributed. So we do not need to apply the normalization step using (18) in this particular example. The result indicates that \( x \) has got its biggest mass of belief on the meta-class \( w_1 \cup w_2 \). In the 5-nearest neighbors of this object, three of them belong to \( w_1 \), whereas the other two nearest neighbors belong to \( w_2 \). This implies that the object is very imprecise with respect to the classes \( w_1 \) and \( w_2 \), and the information we used in the example is not sufficient for making a more specific (refined) classification of this object. So the meta-class \( w_1 \cup w_2 \) can be a good (acceptable) compromise for the classification of this object, which can reduce the risk of the false classification. This solution is also consistent with our intuition. Such fusion result can be considered as a useful source of information to justify the need for other complementary information sources if a more specific decision is absolutely necessary for the problem under consideration.

### 4.3. Expressive power of BK-NN

The expressive power of a method can be seen as its ability to identify and manipulate complex propositions. Let us examine and compare the expressivity power of BK-NN with respect to EK-NN and K-NN through the following simple example.
**Example 3.** Let us consider the simple frame of classes $\Omega = \{w_0, w_1, w_2\}$, and an object $x$, with 2-nearest neighbors respectively labeled by classes $w_1$ and $w_2$ with the corresponding bba’s satisfying BK-NN bba modeling. The set of focal elements of $m_1(\cdot)$ is $F_{12}^{BN-NN} = \{w_1, w_2, \Omega\}$ and the set of focal elements of $m_2(\cdot)$ is $F_{22}^{BN-NN} = \{w_2, \Omega\}$. Therefore, the set of focal elements resulting from the combination rule (17) is

$$F_{22}^{BN-NN} = \{w_1, w_2, w_1 \cup w_2, w_1 \cup w_2, \Omega\}$$

So we see that BK-NN generates seven focal elements (specific classes and meta-classes). With EK-NN method, only three focal elements will be expressed because one would have $F_{12}^{EK-NN} = \{w_1, \Omega\}$ and $F_{22}^{EK-NN} = \{w_2, \Omega\}$ and therefore $F_{12}^{EK-NN} = \{w_1, w_2, \Omega\}$ with DS rule. Clearly, BK-NN bba modeling has a better expressivity power than in EK-NN.

If we consider a larger frame, say $\Omega = \{w_0, w_1, w_2, w_3\}$ with three nearest neighbors with classes $w_1$, $w_2$, and $w_3$, then we will get now a set of 15 focal elements

$$F_{12}^{BN-NN} = \{w_1, w_2, w_3, w_1 \cup w_2, w_1 \cup w_3, w_2 \cup w_3, w_1 \cup w_2 \cup w_3, w_1 \cup w_2 \cup w_3, w_1 \cup w_2 \cup w_3, \Omega\}$$

As we can see, BK-NN can discriminate 15 elements (the specific classes and meta-classes) from a frame of discernment having three classes, whereas K-NN discriminates only three classes and meta-classes. With EK-NN method, only three focal elements will be considered as default value. In the artificial numerical parameters of EK-NN were automatically optimized using the cross validation method is used to select the best parameters. We have determined by the average distances $g(w_i)$ and therefore $\eta = 0.2$ has been chosen as the default value for the selection of meta-classes, and the different value of $\eta$ is applied to show its influence on the results.

In order to show the ability of BK-NN to deal with imprecise information, the classification of the objects was done on the basis that the class should receive the maximal mass of belief.

5. Applications of BK-NN

The effectiveness of BK-NN with respect to EK-NN and K-NN will be shown through three experiments in this section. The parameters of EK-NN were automatically optimized using the method introduced in [31].

For BK-NN, several heuristics have been tested to choose the tuning parameters, and it generally produces good results when acceptance and rejection thresholds are $\gamma_a = 2$ and $\gamma_r = 5$, which can be considered as default value. In the artificial numerical examples of Experiments 1 and 2, the default value of $\gamma_a$ and $\gamma_r$ are applied. In applications of the real data set as in Experiment 3, the cross validation method is used to select the best parameters.

The parameter $\lambda$, was determined by the average distances according to (13). $\eta = 0.2$ has been chosen as the default value for the selection of meta-classes, and the different value of $\eta$ is applied to show its influence on the results.

In order to show the ability of BK-NN to deal with imprecise information, the classification of the objects was done on the basis that the class should receive the maximal mass of belief.

5.1. Experiment 1

Two particular tests have been done in this experiment to explicitly illustrate the difference between EK-NN, K-NN and BK-NN.

(1) Test 1: Two classes of artificial data sets $w_1$ and $w_2$ have been generated from the two 2D-Gaussian distributions $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ characterized by the following mean vectors and covariance matrices:

$$\mu_1 = (0, 0) \quad \text{and} \quad \Sigma_1 = 2 \times I$$

$$\mu_2 = (5, 0) \quad \text{and} \quad \Sigma_2 = 2 \times I$$

where $I$ is the identity matrix of dimension $2 \times 2$.

Each class has 50 training data points and 50 test points. We used $K=7$ nearest neighbors in this test. For notation conciseness, we have denoted $w_{\text{tr}} \triangleq w_{\text{tr}}^\text{test}$, $w_{\text{tr}} \triangleq w_{\text{tr}}^\text{training}$ and $w_{\text{tr}}^{-k} \triangleq w_1 \cup \cdots \cup w_k$.

As we can see in Fig. 4(a), the data sets of $w_1$ and $w_2$ are close to each other with partial overlap. K-NN and EK-NN just commit the objects to two specific classes as shown in Fig. 4(b) and (c). BK-NN provides much more enlarged classifications as shown in Fig. 4(d) than EK-NN and K-NN. With BK-NN, the seven objects labeled by black plus symbol in the middle of the two classes are committed to the meta-class $w_1 \cup w_2$, since the two classes are not very distinguishable for these objects. The meta-class $w_1 \cup w_2$ is in fact a more suitable classification for these seven objects. Two objects quite far from its neighbors are reasonably considered as outliers $w_1 = w_1 \cup w_2$ labeled by black pentagram. One point labeled by a blue triangle is far from the class $w_1$, but not close enough to its neighbors in $w_2$. So this object is considered belonging to the rejection class $w_{\text{tr}}^{-2}$. Similarly, one point far from $w_2$ but not close to $w_1$ is committed to $w_{\text{tr}}^{-1}$ labeled by red triangle. This example shows the expressive power of hyper-credal classification on which the BK-NN is based.

(2) Test 2: In this example, the data set is generated from three 2D Gaussian distributions characterizing the classes $w_1$, $w_2$, and $w_3$ with the following mean vectors and covariance matrices (I being the $3 \times 3$ identity matrix):

$$\mu_1 = (0,0) \quad \text{and} \quad \Sigma_1 = 1.5 \times I$$

$$\mu_2 = (4,0) \quad \text{and} \quad \Sigma_2 = 1.5 \times I$$

$$\mu_3 = (2,4) \quad \text{and} \quad \Sigma_3 = 2 \times I$$

Each class has 50 training data points and 50 test points. We used $K=9$ nearest neighbors in this test.

In Fig. 5, we can see that the three close data sets are classified into three distinct classes $w_1$, $w_2$ or $w_3$ by K-NN and EK-NN. With BK-NN, few objects in the middle of different classes are committed to the meta-classes which is more reasonable. For example, six objects labeled by black plus symbol in middle of $w_1$ and $w_2$ are committed to $w_1 \cup w_2$, since the number of their KNNs in $w_1$ and $w_2$ is close. Similarly for the objects classified into $w_1 \cup w_2$ and $w_1 \cup w_2$, two points labeled by black pentagram are far from the neighbors, and are classified as outliers by BK-NN. This example shows how BK-NN works and its ability to deal with the meta-classes.

5.2. Experiment 2

In this second experiment, we compare the performances of BK-NN with respect to the performances of EK-NN and K-NN on the example given in [6]. The data set is generated from three 3D Gaussian distributions characterizing the classes $w_1$, $w_2$, and $w_3$ with the following mean vectors and covariance matrices (I being the $3 \times 3$ identity matrix):

$$\mu_1 = (1,1,1) \quad \text{and} \quad \Sigma_1 = I$$

$$\mu_2 = (-1,1,0) \quad \text{and} \quad \Sigma_2 = I$$

$$\mu_3 = (0,-1,1) \quad \text{and} \quad \Sigma_3 = 2 \times I$$

Here, we all the partial conflicting beliefs are preserved for the meta-classes.
There are $3 \times 100$ test samples, and the training sets contain $3 \times N$ samples (for $N=100, 200$). Values of $K$ ranging from 5 to 25 neighbors have been investigated. For each pair $(N, K)$, the error rates and imprecision rates reported in Fig. 6 have been averaged on a Monte Carlo simulation based on 10 runs. Moreover, the values $\eta = 0.15, 0.2$ have been tested to evaluate their influences on the results. The mean of the classification error and imprecision rates with different number of training samples (for $N=100,200$ and $\eta = 0.15,0.2$) are presented in Fig. 6(a)–(d).
An object truly originated from \( w_i \) generates an error of classification if it is classified into \( A \) when \( w_i \cap A = \emptyset \). If one gets \( w_i \cap A \neq \emptyset \) and \( A \neq w_i \), then one has an imprecise classification. The error rate \( re \) is calculated by \( re = N_e / T \), where \( N_e \) is the number of objects wrongly classified, and \( T \) is the total number of the objects tested. The imprecision rate \( ri_j \) is calculated by \( ri_j = N_i / T \), where \( N_i \) is the number of objects committed to the meta-classes with cardinality \( j \), and \( T \) is the total number of the objects tested.

As we can see in Fig. 6, K-NN has the biggest error rate, and BK-NN and EK-NN provide better performances than K-NN since the distance information is taken into account in these methods. BK-NN produces the smallest error of classification, because some samples which are difficult to classify into a specific class are reasonably committed to the meta-classes or to rejection classes. The error rate of BK-NN can decrease when \( \eta \) is reasonably committed to the meta-classes or to rejection classes.

The error rate with (a) \( N = 100, \eta = 0.15 \), (b) \( N = 200, \eta = 0.15 \), (c) \( N = 100, \eta = 0.2 \), (d) \( N = 200, \eta = 0.2 \).

5.3. Experiment 3

In this third experiment, four well known real data sets (i.e. Iris data, Ecoli, Wine and Optical Recognition of Handwritten Digits) from the UCI repository [12] are used to evaluate the performance of BK-NN. For Ecoli data set, the three classes named as \( cp \), \( im \) and \( imU \) are selected in the whole Ecoli data set, since these three classes are close and hard to classify. For Handwritten digits data set, we also choose three classes as 1, 4 and 7, since these three handwritten digits could be similar and difficult to correctly recognize in some cases.

The main characteristics of the four data sets are summarized in Table 1, and the detailed information can be found in [12] (http://archive.ics.uci.edu/ml/).

The \( k \)-fold cross validation is performed on the Iris, Ecoli and Digits data sets by different classification methods. In general \( k \) remains an unfixted parameter [15]. We use the simplest 2-fold cross validation here, since it has the advantage that the training and test sets are both large, and each sample is used for both training and testing on each fold. The samples in each classes are randomly assigned to two sets \( S_1 \) and \( S_2 \), and the two sets have equal size. We then train on \( S_1 \) and test on \( S_2 \), followed by training on \( S_2 \) and testing on \( S_1 \). For Digits data set, the test data and training data are separated in UCI repository, and the test data are directly classified using the training data in this experiment. In the training data sets, we use the leave-one-out cross validation method to optimize the tuning parameters \( \gamma_{m} \) and \( \gamma_{c} \) in BK-NN. The best parameters corresponding to the lowest error rate will be found by the grid search with value of step length as 0.1, and the search scope is recommended as \( \gamma_{m} \in [1, 3] \) and \( \gamma_{c} \in [2\gamma_{m}, 3\gamma_{m}] \).

### Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Classes</th>
<th>Attributes</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>3</td>
<td>4</td>
<td>150</td>
</tr>
<tr>
<td>Ecoli</td>
<td>3</td>
<td>7</td>
<td>255</td>
</tr>
<tr>
<td>Wine</td>
<td>3</td>
<td>13</td>
<td>178</td>
</tr>
<tr>
<td>Digits</td>
<td>3</td>
<td>64</td>
<td>1705</td>
</tr>
</tbody>
</table>
meta-classes. When \( Z \) correctly classify are reasonably committed to the associated lowest error rate of classification since some samples hard to required. will be necessary to get better specific results when absolutely meta-classes, and other complementary information sources for doing a correct specific classification of the objects in the meta-classes, and other complementary information sources will be necessary to get better specific results when absolutely required.

The classification results of different methods for iris data (in %).

<table>
<thead>
<tr>
<th>( K )</th>
<th>K-NN ( re )</th>
<th>EK-NN ( re )</th>
<th>BK-NN (( \eta = 0.2 )) ( (re, r_i) )</th>
<th>BK-NN (( \eta = 0.25 )) ( (re, r_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6.00</td>
<td>6.00</td>
<td>(2.00, 6.67)</td>
<td>(2.00, 6.67)</td>
</tr>
<tr>
<td>6</td>
<td>5.53</td>
<td>4.67</td>
<td>(4.00, 2.67)</td>
<td>(4.00, 2.67)</td>
</tr>
<tr>
<td>7</td>
<td>6.00</td>
<td>5.33</td>
<td>(2.00, 5.33)</td>
<td>(2.00, 5.33)</td>
</tr>
<tr>
<td>8</td>
<td>4.67</td>
<td>4.67</td>
<td>(1.33, 6.00)</td>
<td>(1.33, 6.00)</td>
</tr>
<tr>
<td>9</td>
<td>4.67</td>
<td>4.00</td>
<td>(0.67, 8.67)</td>
<td>(0.67, 8.67)</td>
</tr>
<tr>
<td>10</td>
<td>6.00</td>
<td>6.67</td>
<td>(2.00, 6.67)</td>
<td>(2.00, 6.67)</td>
</tr>
<tr>
<td>11</td>
<td>5.33</td>
<td>4.67</td>
<td>(3.33, 4.67)</td>
<td>(3.33, 4.67)</td>
</tr>
<tr>
<td>12</td>
<td>5.33</td>
<td>4.67</td>
<td>(1.33, 8.67)</td>
<td>(1.33, 8.67)</td>
</tr>
<tr>
<td>13</td>
<td>5.33</td>
<td>3.33</td>
<td>(2.67, 4.67)</td>
<td>(1.33, 10.67)</td>
</tr>
<tr>
<td>14</td>
<td>6.00</td>
<td>4.67</td>
<td>(2.00, 8.00)</td>
<td>(2.00, 8.00)</td>
</tr>
<tr>
<td>15</td>
<td>6.67</td>
<td>3.33</td>
<td>(2.00, 8.67)</td>
<td>(2.00, 8.67)</td>
</tr>
</tbody>
</table>

The classification results of different methods for Ecoli data (in %).

<table>
<thead>
<tr>
<th>( K )</th>
<th>K-NN ( re )</th>
<th>EK-NN ( re )</th>
<th>BK-NN (( \eta = 0.2 )) ( (re, r_i) )</th>
<th>BK-NN (( \eta = 0.25 )) ( (re, r_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9.81</td>
<td>9.41</td>
<td>(3.53, 16.05)</td>
<td>(3.53, 16.05)</td>
</tr>
<tr>
<td>6</td>
<td>11.37</td>
<td>10.59</td>
<td>(5.89, 5.87)</td>
<td>(5.89, 5.87)</td>
</tr>
<tr>
<td>7</td>
<td>10.59</td>
<td>9.81</td>
<td>(5.10, 10.18)</td>
<td>(5.10, 10.18)</td>
</tr>
<tr>
<td>8</td>
<td>12.94</td>
<td>10.98</td>
<td>(7.45, 6.66)</td>
<td>(4.71, 16.46)</td>
</tr>
<tr>
<td>9</td>
<td>11.37</td>
<td>11.76</td>
<td>(6.27, 9.79)</td>
<td>(6.27, 10.58)</td>
</tr>
<tr>
<td>10</td>
<td>12.94</td>
<td>12.16</td>
<td>(5.10, 12.93)</td>
<td>(5.10, 12.93)</td>
</tr>
<tr>
<td>11</td>
<td>13.33</td>
<td>12.94</td>
<td>(7.06, 9.79)</td>
<td>(7.06, 9.79)</td>
</tr>
<tr>
<td>12</td>
<td>13.72</td>
<td>11.76</td>
<td>(5.49, 13.32)</td>
<td>(5.49, 13.71)</td>
</tr>
<tr>
<td>13</td>
<td>13.33</td>
<td>12.15</td>
<td>(6.96, 9.41)</td>
<td>(4.32, 17.64)</td>
</tr>
<tr>
<td>14</td>
<td>13.73</td>
<td>12.16</td>
<td>(5.49, 13.33)</td>
<td>(5.49, 14.11)</td>
</tr>
<tr>
<td>15</td>
<td>14.11</td>
<td>12.16</td>
<td>(4.71, 15.68)</td>
<td>(4.71, 15.68)</td>
</tr>
</tbody>
</table>

The classification results of different methods for Wine data (in %).

<table>
<thead>
<tr>
<th>( K )</th>
<th>K-NN ( re )</th>
<th>EK-NN ( re )</th>
<th>BK-NN (( \eta = 0.2 )) ( (re, r_i) )</th>
<th>BK-NN (( \eta = 0.25 )) ( (re, r_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>28.09</td>
<td>28.09</td>
<td>(11.80, 36.52)</td>
<td>(11.80, 36.52)</td>
</tr>
<tr>
<td>6</td>
<td>27.53</td>
<td>25.84</td>
<td>(19.66, 27.53)</td>
<td>(19.66, 27.53)</td>
</tr>
<tr>
<td>7</td>
<td>28.65</td>
<td>27.53</td>
<td>(12.36, 31.46)</td>
<td>(12.36, 31.46)</td>
</tr>
<tr>
<td>8</td>
<td>31.46</td>
<td>30.90</td>
<td>(17.42, 15.73)</td>
<td>(9.55, 38.76)</td>
</tr>
<tr>
<td>9</td>
<td>26.97</td>
<td>25.28</td>
<td>(14.04, 25.84)</td>
<td>(12.92, 29.78)</td>
</tr>
<tr>
<td>10</td>
<td>28.09</td>
<td>26.40</td>
<td>(11.80, 33.71)</td>
<td>(11.80, 33.71)</td>
</tr>
<tr>
<td>11</td>
<td>26.97</td>
<td>25.84</td>
<td>(16.29, 24.16)</td>
<td>(16.29, 24.16)</td>
</tr>
<tr>
<td>12</td>
<td>28.09</td>
<td>27.53</td>
<td>(15.73, 29.21)</td>
<td>(11.24, 35.96)</td>
</tr>
<tr>
<td>13</td>
<td>26.40</td>
<td>26.40</td>
<td>(16.85, 20.79)</td>
<td>(11.24, 41.01)</td>
</tr>
<tr>
<td>14</td>
<td>26.97</td>
<td>28.09</td>
<td>(13.48, 29.21)</td>
<td>(12.36, 32.02)</td>
</tr>
<tr>
<td>15</td>
<td>25.84</td>
<td>26.97</td>
<td>(11.80, 33.71)</td>
<td>(11.80, 33.71)</td>
</tr>
</tbody>
</table>

6. Conclusion

A new belief-based K-nearest neighbor (BK-NN) method has been developed for the classification of the close data sets, and it can reduce the misclassification error by decreasing a bit the specificity of the result. BK-NN is based on the construction of bba's using the acceptance and rejection thresholds related with the distance between the object and its neighbors. These bba's are combined by a new fusion rule which is a compromise between Dempster–Shafer rule and Dubois–Prade rule. The BK-NN offers an enhanced expressive power with respect to K-NN and EK-NN by allowing the hyper-creal classifications on specific classes, on the meta-classes and on the rejection classes as well. The output of BK-NN can be used as a useful mean to indicate when complementary information sources are needed to get more precise classification result in some particular applications, like those involving defense and security. Several experiments based on simulated and real data sets presented in this paper have shown that BK-NN provides a deeper insight into the data, and reduces the misclassification errors by introducing the rejection classes and meta-classes.

Acknowledgments

The authors are grateful to the anonymous reviewers for all their remarks which helped us to clarify and improve the quality of this paper. This work has been partially supported by China Natural Science Foundation (No. 61075029, 61135001) and PhD Thesis Innovation Fund from Northwestern Polytechnical University (No.cx201015).

References

Z.-g. Liu et al. / Pattern Recognition 46 (2013) 834–844

Zhun-ga Liu was born in China, 1984. He received the Bachelor and Master degree from Northwestern Polytechnical University in 2007 and 2010. He started the Ph.D course in Northwestern Polytechnical University since September 2009, and he has studied in Telecom Bretagne since August 2010 for one year. His research work focus on belief functions theory and its application in data classification.

Quan Pan was born in China 1961. He received the Bachelor degree in Huazhong University of Science and Technology, and he received the master and doctor degree in Northwestern Polytechnical University (NWPU) in 1991 and 1997. He has been professor since 1998 in NWPU. His current research interests are data classification and information fusion.

Jean Dezert was born in France, 1962. He received the electrical engineering degree from Ecole Francaise de Radioelectrique Electronique and Informatique (EFREI), Paris, in 1985, the D.E.A. degree in 1986 from the University Paris VII and his Ph.D from the University Paris XI, Orsay, in 1990. Since 1993, he is senior research scientist in the Information Modeling and Processing Department (DTIM) at ONERA. His current research interest focus on belief functions theory and its applications, especially for DSmT which has been developed by him and Prof. Smarandache.