

Fuzzy-belief k-nearest neighbor classifier for uncertain data

Zhun-Ga Liu, Quam Pan, Jean Dezert, Grégoire Mercier, Yong Liu

► **To cite this version:**

Zhun-Ga Liu, Quam Pan, Jean Dezert, Grégoire Mercier, Yong Liu. Fuzzy-belief k-nearest neighbor classifier for uncertain data. Fusion 2014: 17th International Conference on Information Fusion, Jul 2014, Salamanca, Spain. pp.1-8. <hal-01070480>

HAL Id: hal-01070480

<https://hal-onera.archives-ouvertes.fr/hal-01070480>

Submitted on 1 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fuzzy-belief K-nearest neighbor classifier for uncertain data

Zhun-ga Liu^{a,b}, Quan Pan^a, Jean Dezert^c, Gregoire Mercier^b, Yong Liu^a

a. School of Automation, Northwestern Polytechnical University, Xi'an, China. Email: liuzhunga@gmail.com

b. Telecom Bretagne, CNRS UMR 6285 Lab-STICC/CID, Brest, France, Email: Gregoire.Mercier@telecom-bretagne.eu

c. ONERA - The French Aerospace Lab, F-91761 Palaiseau, France. Email: jean.dezert@onera.fr

Abstract—Information fusion technique like evidence theory has been widely applied in the data classification to improve the performance of classifier. A new fuzzy-belief K-nearest neighbor (FBK-NN) classifier is proposed based on evidential reasoning for dealing with uncertain data. In FBK-NN, each labeled sample is assigned with a fuzzy membership to each class according to its neighborhood. For each input object to classify, K basic belief assignments (BBA's) are determined from the distances between the object and its K nearest neighbors taking into account the neighbors' memberships. The K BBA's are fused by a new method and the fusion results are used to finally decide the class of the query object. FBK-NN method works with credal classification and discriminate specific classes, meta-classes and ignorant class. Meta-classes are defined by disjunction of several specific classes and they allow to well model the partial imprecision of classification of the objects. The introduction of meta-classes in the classification procedure reduces the misclassification errors. The ignorant class is employed for outliers detections. The effectiveness of FBK-NN is illustrated through several experiments with a comparative analysis with respect to other classical methods.

Index Terms—data classification; evidential reasoning; belief functions; fuzzy membership; K-NN.

I. INTRODUCTION

In the data classification problem, K-nearest neighbor (K-NN) classifier [1] is a well known non-parametric classification method, and it is simple and effective in many applications. In the original voting K-NN method [1], the test sample (object) is classified to the majority class according to its K-nearest neighbors (KNNs)¹ in the training data space, and the object is committed to only one particular class. In the real applications, one never knows with certainty in fact if an input sample belongs to a particular class. A weighted version of K-NN (WK-NN) taking into account the distance between the object and its KNNs has been proposed in [2] to outperform the voting K-NN method. A more general fuzzy K-NN method is given in [3] that assigns fuzzy membership to the labeled sample, and the class of the test sample is decided based on the distance to the sample's KNNs and these KNNs' memberships. Then, the object is allowed to belong to different classes with a fuzzy membership.

In some cases, the given attribute information may be insufficient for making a specific and correct classification

of some objects, and the attribute data in different classes can partly overlap. These objects will be very difficult to classify correctly into a particular class, since several different classes can be undistinguishable for the objects under the given attributes. Moreover, the data set to classify may also contain some noises and outliers in some applications, which makes the classification problem very hard to solve. A method exploiting fuzzy membership functions is not sufficient to model such imprecision of data. Whereas, the evidence theory [4]–[6] also called Dempster-Shafer theory (DST) can well model the uncertain and imprecise information thanks to the belief functions defined on the power-set of the frame of discernment. The belief functions have been already used in many fields, such as data classification [7]–[10], data clustering [11]–[15], and decision-making [16]. An evidential version of K-NN, denoted by EK-NN [7], [17], have been proposed based on DST [4], and it introduces the ignorant class to model the uncertainty. A fuzzy extended version of EK-NN denoted by FEK-NN has been introduced in [18] to handle the more general situation in which each training sample is considered having some degree of membership to each class.

Let us consider the classification of a data set over a class frame $\Omega = \{w_1, \dots, w_c\}$. In the aforementioned evidential methods, only one extra ignorant class denoted by the whole frame Ω is included in the classification procedure. Belief functions work with the power-set of frame denoted by 2^Ω , which contains all the subsets of Ω . However, in 2^Ω , these partial ignorant classes defined by the disjunction of several classes (e.g. $w_i \cup w_j$ or $w_i \cup w_j \cup w_k$, etc), also called meta-class, are not taken into account in the classical evidential methods. The meta-class is very useful and important to explore the imprecision of the classification, and it can also effectively reduce the misclassifications. In many applications, specially those related to defense and security (like in target classification and tracking), it is generally preferable to get a more robust (and eventually less precise) result that could be precisiated later with additional techniques, than to obtain directly with high risk a wrong precise classification from which an erroneous fatal decision would be drawn. In any applications, there is always a compromise to find between the risk of misclassification error and the precision one wants. In our very recent previous work, a belief K -nearest neighbor (BK-NN) classifier [9] has been developed to deal with uncer-

¹In this paper, KNNs denotes the K nearest neighbors of a given object, whereas K-NN denotes the classical K-Nearest Neighbor classification method.

tain data using the meta-class. Nevertheless, the computation complexity of BK-NN is a bit high since many (i.e. three) tuning parameters are involved, and the classification results of BK-NN seem sensitive to the selected K number of the nearest neighbors because the K value is used for the choice of meta-class. These limitations are not convenient for the real engineering applications. Moreover, the class of each training data is considered as specific and certain in BK-NN, and the potential fuzzy membership of the class of training data is not taken into account.

A new fuzzy-belief K-nearest neighbor classifier denoted by FBK-NN is proposed in this work. It considers rigorously all the possible meta-classes in order to characterize the partial imprecision of class of the uncertain data that are hard to correctly classify. In FBK-NN, one assigns a fuzzy membership to each labeled training sample according to the neighborhood of the sample. A new input sample called object is classified based on its KNNs. K basic belief assignments (BBA's) corresponding to the KNNs are constructed using the assigned memberships of the KNNs and the distances between the object and its KNNs. The K BBA's will be fused using a new proposed method. The class of the sample is determined based on the global fusion results. In FBK-NN, only one tuning parameter related to the selection of meta-class is included, and it can be easily optimized using the training data. The meta-class is selected according to the pignistic probability transformation $BetP(\cdot)$ [5], and it makes the classification results of FBK-NN more robust to the K number than BK-NN [9]. So FBK-NN is more convenient and efficient than BK-NN for the engineering applications.

FBK-NN allows the objects to belong to not only specific classes but also meta-classes with different masses of beliefs, and such type of classification is called a credal classification. The samples that are simultaneously close to several classes (e.g. w_i and w_j) and impossible to correctly classify will be committed to the meta-class defined by the disjunction (union) of these several classes (e.g. $w_i \cup w_j$). This indicates that the given attribute information is not sufficient for the correct and specific classification of the objects, and some other additional information sources or techniques are necessary if a more precise classification is necessary. The samples that are too far from the others will be naturally considered as outliers, since we can not get useful information with respect to their class. The output of FBK-NN is a global basic belief assignment of the object to the specific classes, the meta-classes and the outlier class. Such output is an interesting resulting information source that can be used separately, or combined with some other complementary information sources for getting a final precise (specific) decision of the class in the multi-source information fusion system (e.g. multi-sensor target identification system).

This paper is organized as follows. After a brief introduction of the fuzzy classification and credal classification in section II, we present the new FBK-NN method in details in the section III. Several experiments are then given in the section IV to show the performance of FBK-NN with respect to the main

classical methods. This paper is concluded in the last section.

II. BACKGROUND KNOWLEDGE

A. Fuzzy classification

Let us consider a set of sample vectors (objects) $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ to be classified on the frame of classes $\Omega = \{w_1, \dots, w_c\}$. The fuzzy classification of these objects $\mathbf{y} \in Y$ will specify the degree of membership of each object in each of c classes, and the degree of membership of \mathbf{y}_i belonging to w_j is $\mu^i(w_j)$. In order to satisfy the mathematical tractability, the sum of an object's memberships in the c classes are generally constrained to be one.

B. Basics of belief functions

Belief Functions (BF) theory [4]–[6] is also referred as Dempster-Shafer theory (DST), or evidence theory. In this theory, one starts with a frame of discernment $\Omega = \{w_1, \dots, w_c\}$ consisting of a finite discrete set of mutually exclusive and exhaustive hypotheses (classes). The power-set of Ω , denoted 2^Ω , is the set of all the subsets of Ω . The singleton class (e.g. w_i) is called a specific class. The disjunctions of several single classes that represent the partial ignorance in 2^Ω (e.g. $w_i \cup w_j$, or $w_i \cup w_j \cup w_k$, etc) are called meta-classes. The whole frame of discernment Ω is called the (full) ignorant class and serve to characterize the noise and outlier class. A basic belief assignment (BBA) is a function $m(\cdot)$ from 2^Ω to $[0, 1]$ satisfying $\sum_{A \in 2^\Omega} m(A) = 1, m(\emptyset) = 0$.

The subsets A of Ω such that $m(A) > 0$ are called the focal elements of $m(\cdot)$. The lower and upper bounds of imprecise probability associated with BBA's correspond to the belief function $Bel(\cdot)$ and the plausibility function $Pl(\cdot)$ [4]. $[Bel(\cdot), Pl(\cdot)]$ is interpreted as the imprecise interval of the unknown probability $P(\cdot)$. A BBA can also be approximated into a probability measure using the pignistic probability transformation $BetP(\cdot)$ [5] for the decision-making support. $BetP(\cdot)$ transformation is mathematically defined by

$$BetP(A) = \sum_{B \in 2^\Omega} \frac{|A \cap B|}{|B|} \cdot m(B). \quad (1)$$

where $|B|$ is the cardinality of the element $B \in 2^\Omega$, which is the number of singleton elements included in B . For example, if $B = w_i \cup w_j$, then $|B| = 2$.

III. FUZZY-BELIEF K-NEAREST NEIGHBOR CLASSIFIER

Let us consider the input samples (objects) $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ to be classified over the frame of the classes $\Omega = \{w_0, w_1, \dots, w_c\}$, and the set of labeled training samples $X = \{\mathbf{x}_1, \dots, \mathbf{x}_z\}$. The element w_0 represents the potential unknown class. It is included in Ω here for the exhaustiveness of the frame, and it is also used to distinguish the ignorant class denoted by Ω discriminating the objects too far from all the training samples and the meta-class $w_1 \cup \dots \cup w_c$ describing the objects lying the overlapped zone of all the singleton classes.

The *credal classification* consists to classify each object using the belief functions framework. It can be defined as n -tuple $M = (\mathbf{m}_1, \dots, \mathbf{m}_n)$, where \mathbf{m}_i is the BBA of the object

$y_i \in Y, i = 1, \dots, n$ associated with the different elements (classes) of the power-set² 2^Ω . We recall that if the frame Ω has $|\Omega| = c$ elements ($c > 1$), the objects can be committed with different memberships only to c distinct single classes if we use a fuzzy classification. Whereas, in a credal classification method, the objects can belong to $2^{|\Omega|} > |\Omega|$ elements (specific classes and meta-classes as well) with different masses of belief.

A. Fuzzy membership of the labeled samples

The local information around the training sample can be useful and interesting to specify the membership value, and it can make the classification more robust to the abnormal labeled samples. For example, if one training sample is labeled with w_1 , but all its neighborhoods in the training data space are labeled with w_2 , then it very plausible that this training sample labeled with w_1 is in fact an abnormal sample that seldom occurs in the class w_1 . If we use the local information (e.g. the neighborhood) to assign each labeled sample a fuzzy membership, it can distribute some membership of the abnormal sample to other classes. The distance measure has been used for determination of the membership in fuzzy c-means clustering [19] method. We propose to define the membership of \mathbf{x}_i based on the distance between \mathbf{x}_i and its KNNs as follows

$$\begin{cases} u^i(w_j) = 0.51 + \frac{\sum_{\mathbf{x}_s | L(\mathbf{x}_s) = w_j} d^{-1}(x_i, x_s)}{\sum_{s=1}^K d^{-1}(x_i, x_s)} \times 0.49, & w_j = L(\mathbf{x}_i) \\ u^i(w_k) = \frac{\sum_{\mathbf{x}_s | L(\mathbf{x}_s) = w_k} d^{-1}(x_i, x_s)}{\sum_{s=1}^K d^{-1}(x_i, x_s)} \times 0.49, & w_k \neq L(\mathbf{x}_i) \end{cases} \quad (2)$$

where $\{\mathbf{x}_s, s = 1, 2, \dots, K\}$ is the set of the KNNs of the training sample \mathbf{x}_i , and $d^{-1}(x_i, x_s) \triangleq 1/|\mathbf{x}_i - \mathbf{x}_s|$ which is the inverse of the Euclidean distance between \mathbf{x}_i and \mathbf{x}_s .

The normalization condition $\sum_{w_j \in \Omega} u^i(w_j) = 1$ holds. In the determination of class membership of each training sample, we globally take into account the information including the known label of samples, the number of neighbors in each class and the distances to the neighbors. If one training sample \mathbf{x}_i is in class w_j , it will naturally take the biggest membership value belonging to w_j . The bigger number of neighbors in class w_j should lead to the higher degree of membership in w_j . Moreover, if one neighbor is closer to the sample, this sample will have a bigger membership in the class label of the neighbor. This new method is expected to assign a better suitable membership for the training samples.

B. Determination of basic belief assignments

As with the classical K-NN method, the KNNs of an object y_s are found at first in the FBK-NN method. If the object y_s to classify is very close to one of its K close neighbors

²The element w_0 is not necessarily considered in the determination of power-set, since the meta-classes involved with w_0 like $w_0 \cup w_i$ are not taken into account in this work.

\mathbf{x}_i , we consider that y_s and \mathbf{x}_i must share the similar class membership. Whereas, if y_s is far from \mathbf{x}_i then \mathbf{x}_i provides only little useful information for the class of y_s , and \mathbf{x}_i plays almost a neutral role in the classification for y_s . In Dempster-Shafer's theory of belief functions, a complete ignorant source of evidence is modeled by the basic belief assignment $m(\Omega) = 1$, where Ω is the whole frame of discernment.

In our context, the information provided by \mathbf{x}_i for the class membership of y_s will be represented by a BBA defined by

$$\begin{cases} m^{s,i}(w_j) = \alpha^{s,i} u^i(w_j), w_j \in \Omega \\ m^{s,i}(\Omega) = 1 - \alpha^{s,i} \end{cases} \quad (3)$$

where $\alpha^{s,i}$ is the confidence measure of the class membership of y_s with respect to the training sample \mathbf{x}_i . $\alpha^{s,i}$ is determined based on the distance between y_s and \mathbf{x}_i , and the bigger distance generally leads to the smaller confidence measure. Thus, $\alpha^{s,i}$ should be a decreasing function of $d_{s,i}$. This confidence measure $\alpha^{s,i} \in (0, 1]$ is defined for $\lambda_{L(\mathbf{x}_i)} > 0$ and $d_{s,i} \geq 0$ by:

$$\alpha^{s,i} = e^{-\lambda_{L(\mathbf{x}_i)} d_{s,i}} \quad (4)$$

One gets $L(\mathbf{x}_i) = w_p$ when the training sample \mathbf{x}_i is labeled by class w_p . For the notation conciseness, one uses $\lambda_p \triangleq \lambda_{L(\mathbf{x}_i)}$ in the sequel. The tuning parameter λ_p must be positive, and its determination depends on the training data and it is done as follows. The average distance between each training sample in class w_p and its KNNs is calculated at first, and the mean value \bar{d}^{w_p} of all the average distances for the training samples in w_p is used to calculate λ_p , and it is defined by:

$$\lambda_p = \frac{1}{\bar{d}^p} \quad (5)$$

with

$$\bar{d}^p = \frac{1}{N_p} \sum_{i=1}^{N_p} \frac{1}{K} \sum_{j=1}^K \|\mathbf{x}_i^p - \mathbf{x}_{i,j}\| = \frac{1}{KN_p} \sum_{i=1}^{N_p} \sum_{j=1}^K \|\mathbf{x}_i^p - \mathbf{x}_{i,j}\| \quad (6)$$

where N_p is the number of training samples in class w_p , $\mathbf{x}_i^p, i = 1, \dots, N_p$ are the training samples in w_p , and $\mathbf{x}_{i,j}, j = 1, \dots, K$ are the K nearest neighbors of \mathbf{x}_i^p in training data space.

C. Fusion of basic belief assignments

The K BBA's will be fused for the classification of the objects, and the decision about class of the objects will depend on the global fusion results. The fusion of BBA's consists of two steps:

- 1) the sub-combination of the BBA's obtained from the neighbors having the same class label;
- 2) the global fusion of these sub-combination results.

If several BBA's are obtained from the neighbors with the same class label (e.g. w_j), the most masses of belief in these BBA's should be focused on the same class w_j . Thus, these BBA's are usually not in high conflict. Moreover, the structure of these BBA's is quite usual and considered as appropriate for using DS rule. For this reason and its simplicity, we use DS

rule here to combine these BBA's. The DS combination rule is defined as: $\forall A, B_i, C_i \subseteq \Omega$

$$m_{DS}^{s,j}(A) = \frac{\sum_{\substack{g(w_j) \\ \bigcap_{i=1}^{g(w_j)} B_i=A}} \prod_{i=1}^{g(w_j)} m^{s,i}(B_i)}{1 - \sum_{\substack{g(w_j) \\ \bigcap_{i=1}^{g(w_j)} C_i=\emptyset}} \prod_{i=1}^{g(w_j)} m^{s,i}(C_i)} \quad (7)$$

where $g(w_j)$ is the number of the neighbors whose given label is w_j , and where $m^{s,i}(\cdot)$ is given in (3). Because of the particular structure of the BBA's $m^{s,i}(\cdot)$ having only singletons and Ω as focal elements, the resulting BBA obtained by DS rule (7) also carries in this particular context the same type of focal elements. $m_{DS}^{s,j}(\cdot)$ is the resulting (combined) basic belief assignment obtained from the object y_s and its neighbors with the label w_j .

In these sub-combination results, the most masses of belief in different BBA's are generally committed to different classes. Hence, these sub-combination results are likely in high conflict. If DS rule is still applied, it may produce very unreasonable results which could lead to wrong decision. The mass of partial conflicting belief (e.g. $m(A \cap B) = m_1(A)m_2(B) + m_1(B)m_2(A)$) produced by the conjunctive combination reflects the level of difficulty for committing the objects to class A or to class B . So it is better to commit these partial conflicting beliefs to the corresponding meta-class (e.g. $A \cup B$), which avoids a misclassification. However, if all the partial conflicting beliefs are transferred to meta-classes, then too many objects will be automatically assigned to meta-classes. This can seriously degrade the overall precision of the classification. To circumvent this problem, we need to find an acceptable compromise between the imprecision degree and the error rate for classification. The imprecision degree r_i is defined by $r_i \triangleq \frac{n_i}{T}$, where n_i is the number of objects in meta-classes, and T is the total number of objects. We propose that the partial conflicting beliefs transfer to the meta-class should be done conditionally according to the current context as explained in the following.

Let us consider that the sub-combination results of these BBA's for the object y_s are given by $m_{DS}^{s,1}, \dots, m_{DS}^{s,h}$, where $m_{DS}^{s,j}$ is the sub-combination result of the BBA's obtained from the neighbors labeled by class w_j . In the transferable belief model (TBM) [5], the pignistic probability $BetP(\cdot)$ defined in (1) can approximate a BBA by a probability measure for hard decision making support. In this work, we use the pignistic probability measure to automatically select the meta-classes if partial conflicting belief transfer has to be done.

To explain how the partial conflicting belief transfer is decided from the context, let us start with the fusion of two pieces of sub-combination results $m_{DS}^{s,i}(\cdot)$ and $m_{DS}^{s,j}(\cdot)$. Since the class w_i takes bigger mass of belief than any other class in the original combined BBA's $m^{s,i}(\cdot)$, one has

$BetP_{m_{DS}^{s,i}}(w_i) = \max(BetP_{m_{DS}^{s,i}}(\cdot))$, similarly for w_j with $m^{s,j}(\cdot)$ and $BetP_{m_{DS}^{s,j}}(w_j)$. Thus, w_i is the most likely class of y_s according to $m_{DS}^{s,i}(\cdot)$, whereas w_j is the most likely class of y_s according to $m_{DS}^{s,j}(\cdot)$.

The absolute value of the difference between $BetP_{m_{DS}^{s,i}}(w_i)$ and $BetP_{m_{DS}^{s,j}}(w_j)$ is denoted by $\kappa_s(w_i, w_j)$, that is

$$\kappa_s(w_i, w_j) = \left| \max_{w_i \in \Omega} BetP_{m_{DS}^{s,i}}(w_i) - \max_{w_j \in \Omega} BetP_{m_{DS}^{s,j}}(w_j) \right| \quad (8)$$

If $\kappa_s(w_i, w_j)$ is very small (according to a given threshold), it indicates that y_s is close to both the classes w_i and w_j according to $m_{DS}^{s,i}(\cdot)$ and $m_{DS}^{s,j}(\cdot)$, and w_i and w_j cannot be clearly distinguished for the classification of y_s . So $\kappa_s(w_i, w_j)$ is used to reflect the power of assignment of the classification of y_s with class w_i and w_j . If $\kappa_s(w_i, w_j)$ is smaller than a given little threshold ϵ , it indicates that the class of y_s is not very distinguishable between w_i and w_j . Then, the meta-class $w_i \cup w_j$ should be selected, and the corresponding conflicting belief $m(w_i \cap w_j)$ must be transferred to the meta class $w_i \cup w_j$. If $\kappa_s(w_i, w_j) > \epsilon$ the classes w_i and w_j are considered discriminable, and the meta-class $w_i \cup w_j$ is not necessarily involved in the classification process, and in such case the conflicting mass will be distributed to the other available focal elements through the normalization procedure.

In the global fusion of $h > 2$ pieces of sub-combination results, we must find $BetP_{m_{DS}^{s,max}}(w_{max}) = \max[BetP_{m_{DS}^{s,1}}(w_1), \dots, BetP_{m_{DS}^{s,h}}(w_h)]$ at first. Then one gets the most likely class w_{max} of y_s in betting sense. If $\kappa(w_{max}, w_i) = |BetP_{m_{DS}^{s,max}}(w_{max}) - BetP_{m_{DS}^{s,i}}(w_i)| < \epsilon$ condition is satisfied, it means that the class w_i is also a very likely label class of y_s . So the object y_s will be (imprecisely) associated with the meta-classes built from the disjunctions of specific classes included in the set ψ_s defined by

$$\psi_s = \{w_i | \kappa_s(w_{max}, w_i) < \epsilon\} \quad (9)$$

In order to consider all the classes in a fair and unbiased manner, all the meta-classes whose cardinality is no bigger than $|\psi_s|$ are selected and taken into account in the global fusion process. The set of the selected meta-classes is denoted by Ψ_s .

A new compound rule of combination inspired by DS rule and Dubois-Prade (DP) rule [20] is proposed here for the FBK-NN approach. The new global fusion rule is mathematically defined:

$$m^s(A) = \begin{cases} \frac{1}{k} \sum_{\substack{c \\ \bigcap_{i=1}^c B_i=A}} \prod_{i=1}^c m_{DS}^{s,i}(B_i), & A \in \Omega \text{ or } A = \Omega \\ \frac{1}{k} \sum_{\substack{|A| \\ \bigcup_{i=1}^{|A|} B_i=A, \\ \bigcap_{i=1}^{|A|} B_i=\emptyset}} \prod_{i=1}^{|A|} m_{DS}^{s,i}(B_i) \prod_{j=|A|+1}^c m_{DS}^{s,j}(\Omega), & A \in \Psi_s \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where

$$k = 1 - \sum_{\substack{q \\ \bigcap_{i=1}^q C_i = \emptyset \\ C_i \in \Omega}} \prod_{i=1}^q m_i^s(C_i) \prod_{j=q+1}^c m_j^s(\Omega), \quad q > |\psi_s|. \quad (11)$$

The coefficient $\frac{1}{k}$ in eq. (10) is the normalization factor. In some real applications, if one knows that there is no outlier in the data set, then the mass of belief of the ignorant class Ω that we use to represent the outlier class can be committed to the conflicting beliefs and redistributed to other focal elements.

This new compound combination rule includes two parts: a conjunctive combination part and a disjunctive combination part. The conjunctive combination is used in the first formula of (10) to calculate the belief committed to the specific classes and to the ignorant class as well, since the degree of assignment of the object to a specific class or to the ignorant class depends on the consensus of all the sources of evidences. The belief on the meta-class reflects the imprecision one has to commit the object since it can in fact belong to several specific classes involved in the meta-class. So the disjunctive combination is applied in the second formula of (10) for computing the belief of the selected meta-classes.

The pseudo-code of the FBK-NN is given in Table I to explicitly show the procedure of this new method.

Table I
FUZZY-BELIEF K-NEAREST NEIGHBOR ALGORITHM

Input:
Training samples: $X = \{\mathbf{x}_1, \dots, \mathbf{x}_z\}$ in \mathbb{R}^p
Objects to classify: $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ in \mathbb{R}^p
Parameters:
K : number of nearest neighbors
$\epsilon > 0$: threshold for imprecision degree
Determination of membership of training sample by (2);
for $i=1$ to n
Select the KNNs of \mathbf{y}_i
Construction of K BBA's by (3);
Fusion of BBA's from neighbors with same label by (7);
Selection of meta-classes according to (8);
Global fusion of these sub-combination results by (10);
end

Guideline for choosing the meta-class threshold ϵ : In the applications, the threshold ϵ of FBK-NN must be tuned according to the number of objects in meta-class. The bigger ϵ generally leads to more objects in meta-classes, which is not efficient for the specific classification of the objects. The smaller ϵ produces fewer objects in meta-class, but it may cause more misclassifications for the imprecise objects. So ϵ can be tuned according to the imprecision degree of the fusion results that one is ready to accept. ϵ can also be optimized by the cross-validation (e.g. leave-one-out) in the training

data space. In the optimization procedure, one can find a fine compromise between the imprecision degree and error rate by a grid search of ϵ with the given value of K nearest neighbors.

Example 1 : Let's consider the following frame of classes $\Omega = \{w_0, w_1, w_2, w_3\}$, and an object \mathbf{y} with 5-nearest neighbors including two neighbors labeled by classes w_1 and two by w_2 and one by w_3 . It is supposed that the fuzzy class memberships of the five neighbors determined using eq. (2) are

$$\begin{aligned} \mu^1(w_1) &= 0.7, & \mu^1(w_2) &= 0.3 \\ \mu^2(w_1) &= 0.8, & \mu^2(w_2) &= 0.2 \\ \mu^3(w_1) &= 0.1, & \mu^3(w_2) &= 0.9 \\ \mu^4(w_1) &= 0.1, & \mu^4(w_2) &= 0.7, & \mu^4(w_3) &= 0.2 \\ \mu^5(w_1) &= 0.2, & \mu^5(w_2) &= 0.2, & \mu^5(w_3) &= 0.6 \end{aligned}$$

The confidence measures of the class membership of the neighbors can be calculated according to the distance between \mathbf{y} and its five neighbors using (4), and they are given by:

$$\alpha^1 = 0.8, \quad \alpha^2 = 0.9, \quad \alpha^3 = 0.7, \quad \alpha^4 = 0.9, \quad \alpha^5 = 0.5$$

Then, the BBA's can be constructed using the membership μ^i and confidence measure α^i by (3)

	w_1	w_2	w_3	Ω
\mathbf{m}^1	0.56	0.24	0	0.20
\mathbf{m}^2	0.72	0.18	0	0.10
\mathbf{m}^3	0.07	0.63	0	0.30
\mathbf{m}^4	0.09	0.63	0.18	0.10
\mathbf{m}^5	0.10	0.10	0.30	0.50

One sees that $m^1(\cdot)$ and $m^2(\cdot)$ supports with more weight the label w_1 , $m^3(\cdot)$ and $m^4(\cdot)$ supports w_2 , and $m^5(\cdot)$ supports the label w_3 . Then the BBA's from the neighbors with the same class label are combined using DS rule eq. (7) as follows:

$$\begin{aligned} m_{DS}^1(\cdot) &= [m^1 \oplus m^2](\cdot) \\ m_{DS}^2(\cdot) &= [m^3 \oplus m^4](\cdot) \\ m_{DS}^3(\cdot) &= m^5(\cdot) \end{aligned}$$

where \oplus is the symbolic notation of DS fusion rule.

The numerical result is

	w_1	w_2	w_3	Ω
\mathbf{m}_{DS}^1	0.8304	0.1421	0	0.0275
\mathbf{m}_{DS}^2	0.0522	0.8392	0.0698	0.0388
\mathbf{m}_{DS}^3	0.1	0.1	0.3	0.5

The largest pignistic probability values obtained with (1) for each BBA $m_{DS}^i(\cdot)$, $i = 1, 2, 3$ are

$$\begin{aligned} BetP_{m_{DS}^1}(w_1) &= 0.8304 + (0.0275/3) \approx 0.8396 \\ BetP_{m_{DS}^2}(w_2) &= 0.8392 + (0.0388/3) \approx 0.8521 \\ BetP_{m_{DS}^3}(w_3) &= 0.3 + (0.5/3) \approx 0.4667 \end{aligned}$$

and the maximum of these pignistic probabilities corresponds to $BetP_{m_{DS}^2}(w_2)$, so that the label w_2 serves as the reference label to compute $\kappa_s(w_i, w_j)$ with (8) for selecting (or not) the meta-classes in an eventual partial conflicting belief

transfer. If one chooses $\epsilon = 0.1$ as the meta-class selection threshold, one gets $|BetP_{m_{DS}^2}(w_2) - BetP_{m_{DS}^1}(w_1)| < \epsilon$ and $|BetP_{m_{DS}^2}(w_2) - BetP_{m_{DS}^3}(w_3)| > \epsilon$. The result of this selection test indicates that w_1 and w_2 are in fact not clearly distinguishable for the class of this object according to the given ϵ value. Thus, $\psi = \{w_1, w_2\}$, and all the meta-classes with cardinality less or equal to $|\psi| = 2$ are selected and $\Psi = \{w_1 \cup w_2, w_1 \cup w_3, w_2 \cup w_3\}$. Then, the sub-combination results: $m_{DS}^1(\cdot)$, $m_{DS}^2(\cdot)$ and $m_{DS}^3(\cdot)$ are fused using the new compound combination rule (10), denoted symbolically \otimes . That is

$$m(\cdot) = [m_{DS}^1 \otimes m_{DS}^2 \otimes m_{DS}^3](\cdot)$$

The numerical values of this global fusion are

$$\begin{aligned} m(w_1) &= 0.0844, & m(w_2) &= 0.1618, \\ m(w_3) &= 0.0034, & m(\Omega) &= 0.0010, \\ m(w_1 \cup w_3) &= 0.0715, & m(w_2 \cup w_3) &= 0.0250, \\ m(w_1 \cup w_2) &= \mathbf{0.6529} \end{aligned}$$

One sees that the meta-class $w_1 \cup w_2$ carries the biggest mass of belief. It indicates that the object \mathbf{y} is likely to belong to w_1 or w_2 , but the two classes are not very distinguishable for the object according to the used attribute information. It is difficult to commit the object correctly to a particular specific class w_1 or w_2 . So the meta-class $w_1 \cup w_2$ can be a good compromise for the classification of this object, which can reduce the error, and this solution is also consistent with our intuition.

IV. EXPERIMENTS

Three experiments have been carried out to test and evaluate the performance of FBK-NN with respect to the recent BK-NN and several other traditional methods including K-NN, FK-NN and EK-NN. The parameters of EK-NN were automatically optimized using the method proposed in [17]. In the applications of FBK-NN and BK-NN with real data, the involved tuning parameters are automatically optimized using training data. In order to show the ability of FBK-NN to deal with the meta-classes, each object is decided according to the maximal mass of belief criterion.

A. Experiment 1

This experiment is to explicitly illustrate the difference between the credal classification obtained by FBK-NN and the classical classification done by K-NN and by FK-NN (the fuzzy K-NN). Two classes of artificial data set w_1 and w_2 are obtained from two uniform distributions as shown by Fig. 1-a. Each class has 150 training samples and 150 test samples, and one more noisy sample (a true outlier belonging to the class w_0) is included in the test samples. The uniform distributions of the samples of the two classes are characterized by the following bounds:

	x-label interval	y-label interval
w_1	(-1.5, 1.5)	(-0.15, 0.15)
w_2	(-0.25, 0.25)	(-2, 2)

A particular value of $K = 11$ is selected here, since it provides good results for all the three methods. The classification results of the test objects by different methods are given by Fig. 1-b–1-d. For notation conciseness, we have denoted $w^{te} \triangleq w^{test}$, $w^{tr} \triangleq w^{training}$ and $w_{i,\dots,k} \triangleq w_i \cup \dots \cup w_k$.

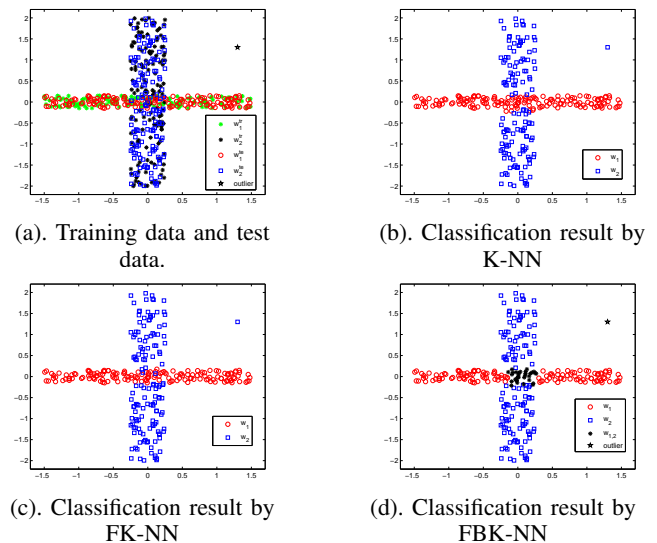


Figure 1. Classification results of 2-class data set by K-NN, EK-NN and FBK-NN .

The objects of classes w_1 and in w_2 are distributed over two overlapping areas following a cross shape as we can see on Fig. 1-a. Naturally the objects belonging to the middle of the cross area are really difficult to associate with a particular class. However, K-NN and FK-NN just commit most of objects of this middle cross area to w_1 as shown on Fig. 1-b and on Fig. 1-c. Obviously, such classification methods generate many misclassification errors. FBK-NN provides one more meta-class $w_1 \cup w_2$ than FK-NN and K-NN as shown on Fig. 1-d. The classes w_1 and w_2 seems undistinguishable for these objects in the intersecting (overlapping) zone. Thus, it is better to prudently assign these objects to the meta-class $w_1 \cup w_2$. By doing this, one reduces the number of misclassification, and one deeply reveals the imprecision degree of class of the objects. An object too far from the others is considered as an outlier by FBK-NN as shown on Fig. 1-d, which means that we cannot get useful information about the class of this object from the data set. However, the outlier is committed with the class w_2 by K-NN and FK-NN due to the inherent limitation of fuzzy classification that cannot model explicitly the outlier class. This example shows the advantaged and interest of the fuzzy-credal classification proposed in the FBK-NN approach.

B. Experiment 2

We consider a 3-class data set composed by three rings as shown on Fig. 2-(a). Each class contains 303 training samples and 303 test objects. The radiuses and centers of the three rings are given by:

	center	radius interval
w_1	(-2, 0)	[3, 4]
w_2	(2, 0)	[3, 4]
w_3	(0, 3.5)	[3, 4]

We selected $K = 11$ nearest neighbors in the three methods. The classification results of test data by K-NN, FK-NN and FBK-NN are respectively shown in Fig. 2-(b)-(d). The meta-class selection threshold $\epsilon = 0.3$ in FBK-NN has been chosen in our simulations.

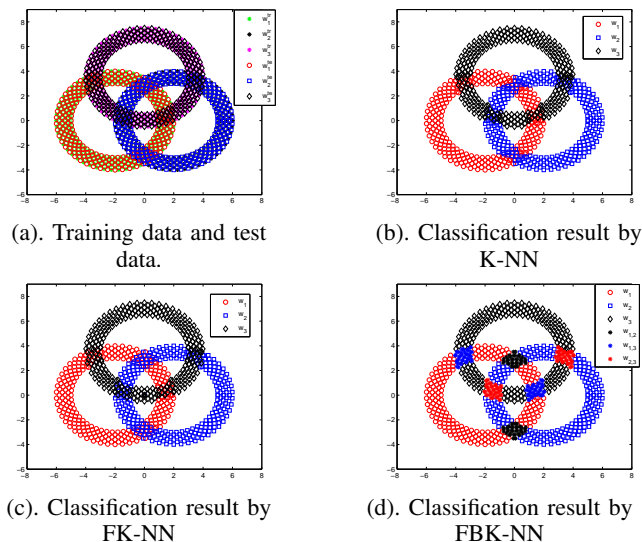


Figure 2. Classification results of a 3-class data set by K-NN, FK-NN and FBK-NN.

We can see that the three rings intersect with each other, and these objects in the overlapped zones are impossible to be correctly classified. In the fuzzy classification of K-NN and FK-NN, all the objects are committed to a particular class as on Fig. 2-(b), (c). In fact, K-NN generates 146 misclassifications, and FK-NN generates 187 misclassifications in this example. In FBK-NN, the objects in the overlapped zones are reasonably considered as belonging to some meta-classes as shown on Fig. 2-(d). This indicates that the used attribute information is not sufficient for making a correct and specific classification of these objects. There are just two misclassifications produced by FBK-NN. In the meanwhile, FBK-NN commits 227 objects in the meta-classes. This example shows the effectiveness of FBK-NN for dealing with the overlapped data in a complex situation by not committing objects to specific classes when there exists a high risk of misclassification.

C. Experiment 3

Four well-known real data sets obtained from UCI Machine Learning Repository [21] (the Glass, Seeds, Ecoli and Breast cancer data sets) are tested in this experiment to evaluate the performance of FBK-NN compared with K-NN, FK-NN, EK-NN and BK-NN. For the Ecoli data set, three classes named as *cp*, *im* and *imU* are selected here, since these three classes

are close and hard to be classified. For the similar reason, we also choose three classes as building-windows-float-processed, building-windows-non-float-processed and vehicle-windows-float-processed from Glass data set. The main characteristics of the four data sets are summarized in Table II, and the detailed information can be found at <http://archive.ics.uci.edu/ml/>.

Table II
BASIC INFORMATION OF THE USED DATA SETS.

name	classes	attributes	instances
Glass	3	10	164
Seeds	3	7	210
Ecoli	3	7	255
Breast cancer	2	9	683

The k -fold cross validation is performed on the four data sets by different classification methods, and k generally remains a free parameter. We use the simplest 2-fold cross validation here, since it has the advantage that the training and test sets are both large. In our simulations, we use the following measures of performances: the misclassification is declared (counted) for one object truly originated from w_i if it is classified into A with $w_i \cap A = \emptyset$. If $w_i \cap A \neq \emptyset$ and $A \neq w_i$ then it will be considered as an imprecise classification. The error rate denoted by Re is calculated by $Re = N_e/T$, where N_e is number of misclassification errors, and T is the number of objects under test. The imprecision rate denoted by Ri_j is calculated by $Ri_j = Ni_j/T$, where Ni_j is number of objects committed to the meta-classes with the cardinality value j . Here we only take Ri_2 since there is no object committed to a meta-class with cardinality value of three.

In the training data sets, we use the leave-one-out cross validation method to optimize the tuning parameters ϵ in FBK-NN and the parameters γ_{t_a} , γ_{t_r} and η in BK-NN. The best parameters corresponding to the suitable (acceptable) compromise between error rate and imprecision rate will be found by the grid search with value of step length as 0.01, on the scope of $\epsilon \in [0, 1]$. The classification results obtained by the different methods with values of K ranging from 5 to 20 are shown in Fig. 3-(a)-(d).

From the Fig. 3-(a)-(d), one sees that FBK-NN and BK-NN provide the smaller error rate than other classical methods, since the objects difficult to classify correctly have been reasonably and automatically committed to the associated meta-classes. The decrease of error rates leads to the increase of imprecision rate, since the objects that are wrongly classified could be assigned into meta-classes. So one should find a compromise between them. It shows that the credal classification can effectively reduce error occurrences, and the meta-classes reveal that the attributes information is not enough to obtain the correct specific class of some objects, and it indicates that other complementary information sources or techniques are necessary if one wants to discriminate the objects in meta-classes. Nevertheless, the classification results of BK-NN seem

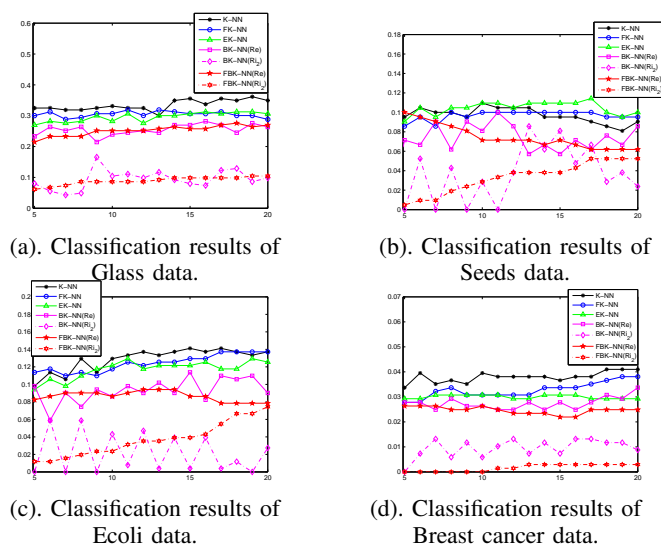


Figure 3. Classification results of real data sets by different methods.

sensitive to the K value, which brings the trouble for choosing K in the application. This is mainly because that K is used in the selection of meta-class in BK-NN. In FBK-NN, $BetP(\cdot)$ is applied for the choice of meta-class. So the results of FBK-NN are not so sensitive to K value as the BK-NN. Moreover, there fewer parameters in FBK-NN to be tuned than in BK-NN. Thus, FBK-NN is more efficient than BK-NN in the real applications.

V. CONCLUSION

A new fuzzy-belief K -nearest neighbor (FBK-NN) method has been presented for classifying the uncertain data based on the fusion of evidence information. The interest and effectiveness of this new classifier have been shown by the comparison with several classical K-NN classifiers and BK-NN using three experiments based on both artificial and real data sets. In FBK-NN, a fuzzy membership is assigned to each labeled training sample to represent its credal assignment to different classes based on its K neighborhood. The K basic belief assignment (BBA) associated with each object to classify are computed from the distances between the object and its KNNs, and the KNNs' membership values. A new compound rule of combination of the K BBA's that manages efficiently the specific classes, the meta classes and the outlier class has been developed and used to classify the object. The object too far from the others will be considered as outliers. As soon as an object (e.g. lying in the overlapped zone of several different classes) is difficult to commit to a specific class, it will be reasonably and automatically committed to a proper meta-class by the FBK-NN. This flexibility provides a deep insight to the data and reduces the misclassification errors.

Acknowledgements

This work has been partially supported by National Natural Science Foundation of China (Nos.61135001, 61374159) and the Fundamental Research Funds for the Central Universities (No. 3102014JCQ01067).

- [1] E. Fix, J.L. Hodges, *Discriminatory analysis. Nonparametric discrimination: consistency properties*, Techn. Rep. 4, Project number 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX, USA, 1951.
- [2] S.A. Dudani, *The distance-weighted k -nearest neighbor rule*, IEEE Trans. on Systems, Man and Cybernetics, SMC-6, pp. 325–327, 1976.
- [3] J.M. Keller, M.R. Gray, J.A. Givens, *A fuzzy K -nearest neighbor algorithm*, IEEE Trans. on Systems, Man and Cybernetics. Vol.15, No. 4, pp. 580–585, 1985.
- [4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [5] P. Smets, *The combination of evidence in the Transferable Belief Model*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, No. 5, pp. 447–458, 1990.
- [6] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, American Research Press, Rehoboth, Vol. 1–3, 2004-2009. <http://fs.gallup.unm.edu/DSMT.htm>
- [7] T. Denœux, *A k -nearest neighbor classification rule based on Dempster-Shafer Theory*, IEEE Trans. on Systems, Man and Cybernetics, Vol. 25, No. 5, pp. 804–813, 1995.
- [8] T. Denœux, P. Smets, *Classification using belief functions: relationship between case-based and model-based approaches*, IEEE Trans. on Systems, Man and Cybernetics, Part B, Vol. 36, No. 6, pp. 1395–1406, 2006.
- [9] Z. Liu, Q. Pan, J. Dezert, *A new belief-based K -nearest neighbor classification method*, Pattern Recognition, Vol. 46, No. 3, pp. 834–844, 2013.
- [10] H. Laanaya, A. Martin, D. Aboutajdine, A. Khenchaf, *Support vector regression of membership functions and belief functions - Application for pattern recognition*, Information Fusion, Vol. 11, No. 4, pp. 338–350, 2010.
- [11] M.-H. Masson, T. Denœux, *ECM: An evidential version of the fuzzy c -means algorithm*, Pattern Recognition, Vol. 41, No. 4, pp. 1384–1397, 2008.
- [12] Z. Liu, J. Dezert, G. Mercier, Q. Pan, *Belief c -means: An extension of fuzzy c -means algorithm in belief functions framework*, Pattern Recognition Letters, Vol. 33, pp. 291–300, 2012.
- [13] Z. Liu, J. Dezert, Q. Pan, Y. Cheng *A new evidential c -means clustering method*, in Proc. of Int. Conf. on Information Fusion (Fusion 2012), Singapore, July 2012.
- [14] T. Denœux, *Maximum likelihood estimation from Uncertain Data in the Belief Function Framework*, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 1, pp. 119–130, 2013.
- [15] V. Antoine, B. Quost, M.-H. Masson, T. Denœux, *CEVCLUS : Evidential clustering with instance-level constraints for relational data*, Soft Computing, DOI: 10.1007/s00500-013-1146-z, 2013.
- [16] Z. Liu, J. Dezert, Q. Pan, G. Mercier, *Combination of sources of evidence with different discounting factors based on a new dissimilarity measure*, Decision Support Systems, Vol. 52, pp. 133–141, 2011.
- [17] L.M. Zouhal, T. Denœux, *An evidence-theoretic k -NN rule with parameter optimization*, IEEE Trans. on Systems, Man and Cybernetics, Part C, Vol. 28, No. 2, pp. 263–271, 1998.
- [18] L.M. Zouhal, T. Denœux, *Generalizing the evidence theoretic k -NN rule to fuzzy pattern recognition*, In Proc. of 2nd Int. ICSC Symposium on Fuzzy Logic and Applications, pp. 294–300, Zurich, Switzerland, 1997.
- [19] J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New-York, 1981.
- [20] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Computational Intelligence, Vol. 4, No. 4, pp. 244–264, 1988.
- [21] A. Frank, A. Asuncion, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, Irvine, CA, USA, 2010.