Evidential Editing K-Nearest Neighbor Classifier

Lianmeng Jiao[1,2], Thierry Denœux[1], and Quan Pan[2]

[1] Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc UMR 7253, Compiègne, France
[2] School of Automation, Northwestern Polytechnical University Xi'an, 710072, P.R. China
lianmeng.jiao@utc.fr

Abstract. One of the difficulties that arises when using the K-nearest neighbor rule is that each of the labeled training samples is given equal importance in deciding the class of the query pattern to be classified, regardless of their typicality. In this paper, the theory of belief functions is introduced into the K-nearest neighbor rule to develop an evidential editing version of this algorithm. An evidential editing procedure is proposed to reassign the original training samples with new labels represented by an evidential membership structure. With the introduction of the evidential editing procedure, the uncertainty of noisy patterns or samples in overlapping regions can be well characterized. After the evidential editing, a classification procedure is developed to handle the more general situation in which the edited training samples are assigned dependent evidential labels. Two experiments based on synthetic and real data sets were carried out to show the effectiveness of the proposed method.

Keywords: Data classification, K-nearest neighbor, Theory of belief functions, Evidential editing

1 Introduction

The K-nearest neighbor (KNN) rule, first proposed by Fix and Hodges [6], is one of the most popular and successful pattern classification techniques. Given a set of N labeled training samples $\mathcal{T} = \{(\mathbf{x}^{(1)}, \omega^{(1)}), \cdots, (\mathbf{x}^{(N)}, \omega^{(N)})\}$ with input vectors $\mathbf{x}^{(i)} \in \mathcal{R}^D$ and class labels $\omega^{(i)} \in \{\omega_1, \cdots, \omega_M\}$, the KNN rule classifies a query pattern $\mathbf{y} \in \mathcal{R}^D$ based on the class labels represented by its K nearest neighbors (according to, e.g., the Euclidean distance measure) in the training set $\mathcal{T}$. The basic rationale for the KNN rule is both simple and intuitive: samples close in feature space are likely to belong to the same class. The KNN rule is a suboptimal procedure. However, it has been shown that, in the infinite sample situation, the error rate for the 1-NN rule is bounded above by no more than twice the optimal Bayes error rate. Furthermore, as K increases, this error rate approaches the optimal rate asymptotically [7].

One of the problems encountered in using the KNN classifier is that each of the training samples is considered equally important in the assignment of the class label to the query pattern. This limitation frequently causes difficulty in regions where the data sets from different classes overlap. Atypical samples are given as much weight as those that are truly representatives of the clusters. Furthermore, it may be argued that training samples containing noise should not be given equal weight. In order to overcome this difficulty, the editing procedure was proposed to preprocess the original training samples and the KNN rule was used to classify the query pattern based on the edited training samples [10, 11, 16]. According to the structure of the edited labels, the editing procedures can be divided into two categories: crisp and soft editing. In [16], Wilson proposed a simple editing procedure to preprocess the training set. This procedure classifies a training sample $\mathbf{x}^{(i)}$ using the KNN rule with the remainder of the training set, and deletes it from the original training set if its original label $\omega^{(i)}$ does not agree with the classification result. Later, concerned with the possibility of large amounts of samples being removed from the training set, Koplowitz and Brown [11] developed a modification of the simple editing technique. For a given value of K, another parameter K' is defined such that $(K+1)/2 \leq K' \leq K$. Instead of deleting all the conflicting samples, if a particular class (excluding the original class) has at least K' representatives among these K nearest neighbors, then $\mathbf{x}^{(i)}$ is labeled according to that majority class. Essentially, both the simple editing procedure and its modification belong to the category of crisp editing procedures, in which each edited sample is either removed or assigned to a single class. In order to overcome the difficulty of the crisp editing method in severely noisy conditions, a fuzzy editing procedure was proposed that reassigns fuzzy membership to each training sample $\mathbf{x}^{(i)}$ based on its K nearest neighbors [10]. This fuzzy editing procedure belongs to the soft editing category, in which each edited sample can be assigned to several classes. It provides more detailed information about the samples' membership than the crisp editing procedures.

Different kinds of uncertainty may coexist in real-world classification problems, e.g., fuzziness may coexist with imprecision or incompleteness. The fuzzy editing procedure, which is based on fuzzy set theory [17], cannot address imprecise or incomplete information effectively in the modeling and reasoning processes. In contrast, the theory of belief functions [1, 14, 15], also called Dempster-Shafer theory, can well model imprecise or incomplete information thanks to the belief functions defined on the power set of the frame of discernment. The theory of belief functions has already been used in the pattern classification field [2, 4, 8, 9, 12]. An evidential version of KNN, denoted by EKNN [2], has been proposed based on the theory of belief functions; it introduces the ignorance class to model the uncertainty. In [12], the EKNN was further extended to deal with uncertainty using a meta-class. Neither the EKNN method nor its extension consider any editing procedure and the original training set is used to make classification. More recently, an editing procedure for multi-label classification was developed in [9] based on an evidential multi-label KNN rule (EMLKNN) [5], but it essentially belongs to the crisp editing category as each edited sample is either removed or assigned to a new set of classes without considering the class membership degrees.

In this paper, an evidential editing K-nearest neighbor (EEKNN) is proposed based on the theory of belief functions. The proposed EEKNN classifier contains two stages: evidential editing and classification. First, an evidential editing procedure reassigns the original training samples with new labels represented by an evidential membership structure. Compared with the fuzzy membership used in fuzzy editing, the evidential labels provide more expressiveness to characterize the imprecision for those samples with great noise or in overlapping regions. For a training sample $\mathbf{x}^{(i)}$, if there is no imprecision among the frame of discernment, the evidential membership reduces to the fuzzy membership. After the evidential editing procedure, a classification procedure is developed to handle the more general situation in which the edited training samples are assigned dependent evidential labels.

The rest of this paper is organized as follows. In Section 2, the basics of belief function theory are recalled. The evidential editing K-nearest neighbor (EEKNN) classifier is developed in Section 3 and then two experiments are developed to evaluate the performance of the proposed EEKNN in Section 4. Finally, Section 5 concludes the paper.

2 Background on the Theory of Belief Functions

In the theory of belief functions [1, 14], a problem domain is represented by a finite set $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$ of mutually exclusive and exhaustive hypotheses called the *frame of discernment*. A *basic belief assignment* (BBA) expressing the belief committed to the elements of 2^Θ by a given source of evidence is a mapping function m: $2^\Theta \rightarrow [0, 1]$, such that

$$\mathrm{m}(\emptyset) = 0 \;\text{ and }\; \sum_{A \in 2^\Theta} \mathrm{m}(A) = 1. \tag{1}$$

Elements $A \in 2^\Theta$ having $\mathrm{m}(A) > 0$ are called the *focal elements* of the BBA m. Each number $\mathrm{m}(A)$ measures the degree of belief exactly assigned to a proposition A. The belief assigned to Θ, is referred to as the degree of *global ignorance*. A BBA is said to be *simple* if it has the following form

$$\begin{cases} \mathrm{m}(A) = 1 - w \\ \mathrm{m}(\Theta) = w, \end{cases} \tag{2}$$

for some $A \subset \Theta$ and $w \in [0, 1]$. Let us denote such a mass function as A^w.

Shafer [14] also defines the *belief* and *plausibility functions* as follows

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} \mathrm{m}(B) \;\text{ and }\; \mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} \mathrm{m}(B), \;\text{ for all } A \in 2^\Theta. \tag{3}$$

$\mathrm{Bel}(A)$ represents the exact support to A and its subsets, and $\mathrm{Pl}(A)$ represents the total possible support to A and its subsets. The interval $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$ can be seen as the lower and upper bounds of support to A.

For decision making, Smets [15] proposed the *pignistic probability* BetP to approximate the unknown probability in $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$, given by

$$\mathrm{BetP}(A) = \sum_{B \cap A \neq \emptyset} \frac{|A \cap B|}{|B|} \mathrm{m}(B), \;\text{ for all } A \in 2^\Theta, \tag{4}$$

where $|X|$ is the cardinality of set X.

Two useful operations in the manipulation of belief functions are *Shafer's discounting operation* and *Dempster's rule of combination*. The discounting operation is used when a source of evidence provides a BBA m, but one knows that this source has a probability of $\alpha \in [0,1]$ being reliable. Then, one may adopt $(1-\alpha)$ as the discount rate, which results in a new BBA ${}^{\alpha}\mathrm{m}$ defined by

$$ {}^{\alpha}\mathrm{m}(A) = \begin{cases} \alpha \mathrm{m}(A), & \text{for } A \neq \Theta \\ \alpha \mathrm{m}(\Theta) + (1-\alpha), & \text{for } A = \Theta. \end{cases} \tag{5} $$

Several distinct bodies of evidence characterized by different BBAs can be combined using Dempster's rule. Mathematically, the combination of two BBAs m_1 and m_2 defined on the same frame of discernment Θ yields the following BBA,

$$ \mathrm{m}_1 \oplus \mathrm{m}_2 = \begin{cases} 0, & \text{for } A = \emptyset \\ \frac{\sum_{B\cap C=A} \mathrm{m}_1(B)\mathrm{m}_2(C)}{1-\sum_{B\cap C=\emptyset} \mathrm{m}_1(B)\mathrm{m}_2(C)}, & \text{for } A \in 2^{\Theta} \text{ and } A \neq \emptyset. \end{cases} \tag{6} $$

To combine BBAs induced by nondistinct bodies of evidence, a *cautious rule of combination* and, more generally, a family of parameterized *t-norm based combination rules* with behavior ranging between Dempster's rule and the cautious rule are proposed in [3]:

$$ \mathrm{m}_1 \circledast_s \mathrm{m}_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w_1(A) \top_s w_2(A)}, \tag{7} $$

where m_1 and m_2 are *separable* BBAs [14], such that $\mathrm{m}_1 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w_1(A)}$ and $\mathrm{m}_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w_2(A)}$, and $\top_s$ is the Frank's parametrized family of t-norms:

$$ a \top_s b = \begin{cases} a \wedge b, & \text{if } s = 0 \\ ab, & \text{if } s = 1 \\ \log_s \left(1 + \frac{(s^a-1)(s^b-1)}{s-1} \right), & \text{otherwise,} \end{cases} \tag{8} $$

for all $a, b \in [0,1]$, where s is a positive parameter. When $s = 0$, the t-norm based rule corresponds to cautious rule and when $s = 1$, it corresponds to Dempster's rule.

3 Evidential Editing K-Nearest Neighbor Classifier

Let us consider an M-class classification problem and let $\Omega = \{\omega_1, \cdots, \omega_M\}$ be the set of classes. Assuming that a set of N labeled training samples $\mathcal{T} = \{(\mathbf{x}^{(1)}, \omega^{(1)}), \cdots, (\mathbf{x}^{(N)}, \omega^{(N)})\}$ with input vectors $\mathbf{x}^{(i)} \in \mathcal{R}^D$ and class labels $\omega^{(i)} \in \Omega$ are available, the problem is to classify a query pattern $\mathbf{y} \in \mathcal{R}^D$ based on the training set $\mathcal{T}$.

The proposed evidential editing K-nearest neighbor (EEKNN) procedure is composed of the following two stages:

1. *Preprocessing (evidential editing)*: The evidential editing algorithm assigns evidential labels to each labeled sample.
2. *Classification*: The class of the query pattern is decided based on the distance to the sample's K nearest neighbors and these K nearest neighbors' evidential membership information.

3.1 Evidential Editing

The goal of the evidential editing is to assign each sample in the training set $\mathcal{T}$ with a new soft label with an evidential structure as follows:

$$\mathcal{T}' = \{(\mathbf{x}^{(1)}, \mathrm{m}^{(1)}), (\mathbf{x}^{(2)}, \mathrm{m}^{(2)}), \cdots, (\mathbf{x}^{(N)}, \mathrm{m}^{(N)})\}, \tag{9}$$

where $\mathrm{m}^{(i)}$, $i = 1, 2, \cdots, N$, are BBAs defined on the frame of discernment Ω.

The problem is now to compute an evidential label for each training sample. In [2], an evidential K-nearest neighbor (EKNN) rule was proposed based on the theory of belief functions, where the classification result of the query pattern is a BBA. In the following part, we use the EKNN rule to carry out the evidential editing.

For each training sample $\mathbf{x}^{(i)}$, $i = 1, 2, \cdots, N$, we denote the leave-it-out training set as $\mathcal{T}^{(i)} = \mathcal{T} \setminus \{(\mathbf{x}^{(i)}, \omega^{(i)})\}$, $i = 1, 2, \cdots, N$. Now, we consider the evidential editing for one training sample $\mathbf{x}^{(i)}$ on the basis of the information contained in $\mathcal{T}^{(i)}$. For the training sample $\mathbf{x}^{(i)}$, each neighbor $\mathbf{x}^{(j)}$ $(j \neq i)$ provides an item of evidence regarding the class membership of $\mathbf{x}^{(i)}$ as follows

$$\begin{cases} \mathrm{m}^{(i)}(\{\omega^q\} \mid \mathbf{x}^{(j)}) = \alpha\phi_q(d_{ij}) \\ \mathrm{m}^{(i)}(\Omega \mid \mathbf{x}^{(j)}) \quad\;\; = 1 - \alpha\phi_q(d_{ij}) \\ \mathrm{m}^{(i)}(A \mid \mathbf{x}^{(j)}) \quad\;\; = 0, \;\; \forall A \in 2^{\Omega} \setminus \{\Omega, \{\omega_q\}\}, \end{cases} \tag{10}$$

where $d_{ij} = d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, ω_q is the class label of $\mathbf{x}^{(j)}$ (that is, $\omega^{(j)} = \omega_q$), and α is a parameter such that $0 < \alpha < 1$. As suggested in [2], $\alpha = 0.95$ can be used to obtain good results on average. When d is the Euclidean distance, a good choice for ϕ_q is

$$\phi_q(d) = \exp(-\gamma_q d^2), \tag{11}$$

with γ_q being a positive parameter associated to class ω_q and can be heuristically set to the inverse of the mean squared Euclidean distance between training samples belonging to class ω_q.

Based on the distance $d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, we select the $Kinit$ nearest neighbors of $\mathbf{x}^{(i)}$ in training set $\mathcal{T}^{(i)}$ and calculate the corresponding $Kinit$ BBAs in the above way. As the items of evidence from different neighbors are independent, the $Kinit$ BBAs are combined using Dempster's rule displayed as Eq. (6) to form a resulting BBA $\mathrm{m}^{(i)}$, synthesizing the final evidential membership regarding the label of $\mathbf{x}^{(i)}$ as

$$\mathrm{m}^{(i)} = \mathrm{m}^{(i)}(\cdot \mid \mathbf{x}^{(i_1)}) \oplus \mathrm{m}^{(i)}(\cdot \mid \mathbf{x}^{(i_2)}) \oplus \cdots \oplus \mathrm{m}^{(i)}(\cdot \mid \mathbf{x}^{(i_{Kinit})}), \tag{12}$$

where $i_1, i_2, \cdots, i_{Kinit}$ are the indices of the $Kinit$ nearest neighbors of $\mathbf{x}^{(i)}$ in $\mathcal{T}^{(i)}$.

3.2 Classification

After the evidential editing procedure introduced in Section 3.1, the problem now turns into classifying a query pattern $\mathbf{y} \in \mathcal{R}^D$ based on the new edited training set $\mathcal{T}'$ as shown in Eq. (9). In this section, we extend the evidential K-nearest neighbor (EKNN) rule [2] to handle the more general situation in which the edited training samples are assigned dependent evidential labels. This classification procedure is composed of the following two steps: first, the BBAs from the query pattern's K nearest neighbors are computed; then, the K BBAs are combined to obtain the final result.

Determination of the BBAs Considering the K nearest neighbors of the query pattern $\mathbf{y}$, if one training sample $\mathbf{x}^{(i)}$ is very close to $\mathbf{y}$, it means that $\mathbf{x}^{(i)}$ is a very reliable piece of evidence for the classification of $\mathbf{y}$. In contrast, if $\mathbf{x}^{(i)}$ if far from $\mathbf{y}$, then it provides only little reliable evidence. In the theory of belief functions, Shafer's discounting operation can be used to discount the unreliable evidence before combination.

Denote $\mathrm{m}^{(i)}$ as the class membership of the training sample $\mathbf{x}^{(i)}$, and β_i as the confidence degree of the class membership of $\mathbf{y}$ with respect to the training sample $\mathbf{x}^{(i)}$. The evidence provided by $\mathbf{x}^{(i)}$ for the class membership of $\mathbf{y}$ is represented with a discounted BBA ${}^{\beta_i}\mathrm{m}^{(i)}$ by discounting $\mathrm{m}^{(i)}$ with a discount rate $1-\beta_i$. The confidence degree β_i is determined based on the distance d_i between $\mathbf{x}^{(i)}$ and $\mathbf{y}$, in such a way that a larger distance results in a smaller confidence degree. Thus, β_i should be a decreasing function of d_i. We use a similar decreasing function with Eq. (11) to define the confidence degree $\beta_i \in (0,1]$ as

$$\beta_i = \exp(-\lambda_i d_i^2), \tag{13}$$

where λ_i is a positive parameter associated to the training sample $\mathbf{x}^{(i)}$ and is defined as

$$\lambda_i = \left(\sum_{q=1}^{M} \mathrm{m}^{(i)}(\{\omega_q\})\overline{d}^q + \mathrm{m}^{(i)}(\Omega)\overline{d} \right)^{-2}, \tag{14}$$

with $\overline{d}$ being the mean distance between all training samples, and $\overline{d}^q$ being the mean distance between training samples belonging to each class ω_q, $q=1,2,\cdots,M$.

Combination of the BBAs To make a decision about the class of the query pattern $\mathbf{y}$, the generated K BBAs should be combined to obtain the final fusion result. For combination, Dempster's rule lies in the assumption that the items of evidence combined be distinct or, in other words, that the information sources be independent. However, in the editing process, common training samples may be used for calculating the class membership of different edited samples. Therefore, the items of evidence from different edited samples to classify the query pattern $\mathbf{y}$ cannot be regarded as independent.

To account for this dependence, we use the parameterized t-norm based combination rule shown in Eq. (7) to obtain the final combination result for query pattern $\mathbf{y}$ as

$$\mathrm{m} = {}^{\beta_{i_1}}\mathrm{m}^{(i_1)} \circledast_s {}^{\beta_{i_2}}\mathrm{m}^{(i_2)} \circledast_s \cdots \circledast_s {}^{\beta_{i_K}}\mathrm{m}^{(i_K)}, \tag{15}$$

where $i_1, i_2, \cdots, i_K$ are the indices of the K nearest neighbors of $\mathbf{y}$ in $\mathcal{T}'$. The selection of parameter s depends on the potential dependence degrees of the edited samples. In practice, we can use the cross-validation test to search for the optimal t-norms based combination rule.

For making decisions based on the above combined BBA m, the pignistic probability BetP shown in Eq. (4) is used and the query pattern $\mathbf{y}$ is assigned to the class with the maximum pignistic probability.

4 Experiments

The performance of the proposed evidential editing K-nearest neighbor (EEKNN) classifier was compared with other nearest-neighbor-based methods (the modified simple editing KNN (SEKNN) [11], the fuzzy editing KNN (FEKNN) [10] and the evidential KNN (EKNN) [2]) through two different types of experiments. In the first experiment, the behavior of the proposed method was studied using synthetic data sets. In the second experiment, six real benchmark data sets from the UCI repository [13] were used to compare the methods.

4.1 Synthetic data sets test

This experiment was designed to evaluate the proposed EEKNN with other nearest-neighbor-based methods using synthetic data sets with different class overlapping ratios, defined as the number of training samples in the overlapping region divided by the total number of training samples. A training sample $\mathbf{x}^{(i)}$ is considered to be in the overlapping region if its corresponding maximum plausibility $\text{Pl}_{\max}^{(i)}$ after editing is less than a set upper bound Pl^*, namely, $\text{Pl}^* = 0.9$. A two-dimensional three-class classification problem was considered. The following class-conditional normal distributions were assumed. For comparisons, we changed the variance of each distribution to control the class overlapping ratio.

Case 1 Class A: $\mu_A = (6,6)^T, \Sigma_A = 3\mathbf{I}$; Class B: $\mu_B = (14,6)^T, \Sigma_B = 3\mathbf{I}$; Class C: $\mu_C = (14,14)^T, \Sigma_C = 3\mathbf{I}$. Overlapping ratio $\rho = 6.67\%$
Case 2 Class A: $\mu_A = (6,6)^T, \Sigma_A = 4\mathbf{I}$; Class B: $\mu_B = (14,6)^T, \Sigma_B = 4\mathbf{I}$; Class C: $\mu_C = (14,14)^T, \Sigma_C = 4\mathbf{I}$. Overlapping ratio $\rho = 10.00\%$
Case 3 Class A: $\mu_A = (6,6)^T, \Sigma_A = 5\mathbf{I}$; Class B: $\mu_B = (14,6)^T, \Sigma_B = 5\mathbf{I}$; Class C: $\mu_C = (14,14)^T, \Sigma_C = 5\mathbf{I}$. Overlapping ratio $\rho = 21.33\%$

A training set of 150 samples and a test set of 3000 samples were generated from the above distributions using equal prior probabilities. For each case, 30 trials were performed with 30 independent training sets. Average test classification rates and the corresponding 95% confidence intervals were calculated. For the proposed EEKNN method, the best values for the parameters $Kinit$ and s were determined in the sets $\{3, 6, 9, 12, 15, 18, 21, 24\}$ and $\{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 0\}$, respectively, by cross-validation. For all the considered method, values of K ranging from 1 to 25 have been investigated. Fig. 1 shows the classification results for synthetic data sets with different overlapping ratios. It can be seen that, for the three cases, the EEKNN method provides better classification performance than other nearest-neighbor-based methods. With the increase of the class overlapping ratio, the performance improvement becomes more important. Furthermore, the EEKNN method is not sensitive to the value of K and it performs well even with a small value of K.

4.2 Benchmark data sets test

The main characteristics of the six real data sets used in this experiment are summarized in Table 1. To assess the results, we considered the resampled paired test. A series of

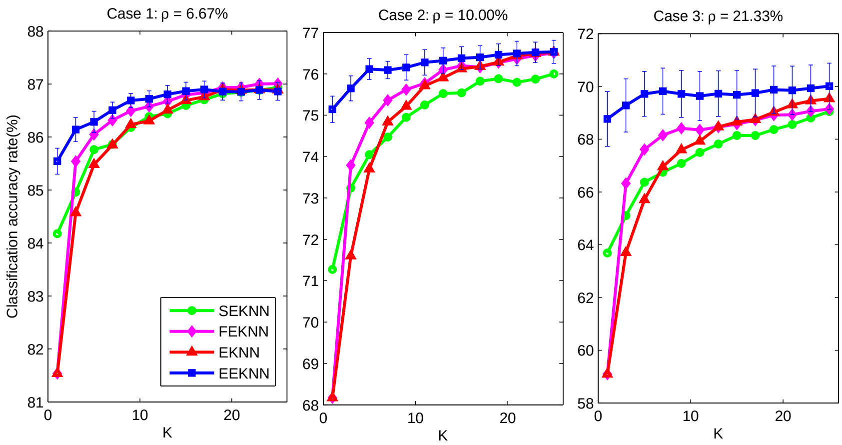


Fig. 1. Classification results for synthetic data sets with different overlapping ratios

30 trials was conducted. In each trials, the available samples were randomly divided into a training set and a test set (with equal sizes). For each data set, we calculated the average classification rates of the 30 trials and the corresponding 95% confidence intervals. For the proposed EEKNN method, the best values for the parameters $Kinit$ and s were determined with the same procedure used in the previous experiment. For all the considered method, values of K ranging from 1 to 25 have been investigated.

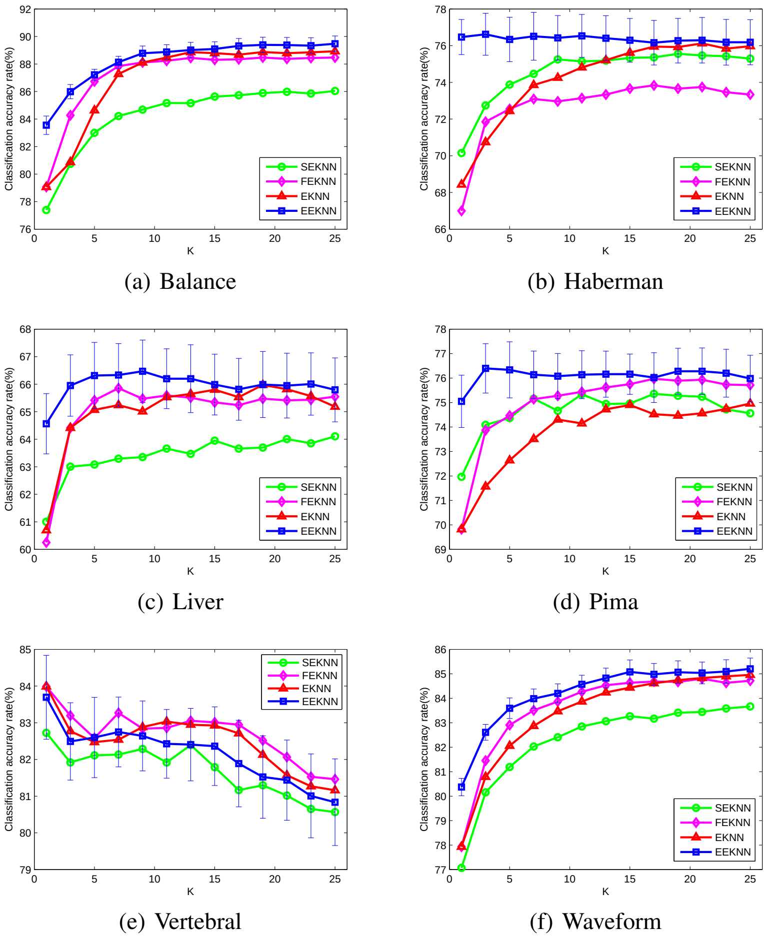
Table 1. Description of the benchmark data sets employed in the study

Data set	# Instances	# Features	# Classes	Overlapping ratio
Balance	625	4	3	19.23%
Haberman	306	3	2	18.59%
Liver	345	6	2	19.19%
Pima	336	8	2	19.05%
Vertebral	310	6	3	11.20%
Waveform	5,000	21	3	19.60%

Fig. 2 shows the classification results of different methods for benchmark data sets. It can be seen that, for data sets with high overlapping ratios, like *Balance*, *Haberman*, *Liver*, *Pima* and *Waveform*, the EEKNN method provides better classification performance than other nearest-neighbor-based methods, especially for small value of K. In contrast, for those data sets with relatively low overlapping ratios, like *Vertebral*, the classification performances of different methods were quite similar. The reason is that, for this data set, the best classification performance was obtained when K took a small value and, under this circumstance, the evidential editing cannot improve the classification performance.

5 Conclusions

An evidential editing K-nearest neighbor (EEKNN) classifier has been developed based on an evidential editing procedure that reassigns the original training samples with new labels represented by an evidential membership structure. Thanks to this procedure, patterns situated in overlapping regions have less influence on the decisions. Our results show that the proposed EEKNN classifier achieves better performance than other considered nearest-neighbor-based methods, especially for data sets with high overlapping ratios. In particular, the proposed EEKNN classifier is not sensitive to the value of K and it can gain a quite good performance even with a small value of K. This is an advantage in time or space-critical applications, in which only a small value of K is permitted in the classification process.



(a) Balance (b) Haberman

(c) Liver (d) Pima

(e) Vertebral (f) Waveform

Fig. 2. Classification results of different methods for benchmark data sets

Acknowledgments. This work was carried out in the framework of the Labex MS2T, which was funded by the French Government, through the program "Investments for the future" managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02).

References

1. Dempster, A.: Upper and lower probabilities induced by multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
2. Denœux, T.: A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 25, 804–813 (1995)
3. Denœux, T.: Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence 172, 234–264 (2008)
4. Denœux, T., Smets, P.: Classification using belief functions relationship between case-based and model-based approaches. IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics 36, 1395–1406 (2006)
5. Denœux, T., Younes, Z., Abdallah, F.: Representing uncertainty on set-valued variables using belief functions. Artificial Intelligence 174, 479–499 (2010)
6. Fix, E., Hodges, J.: Discriminatory analysis, nonparametric discrimination: consistency properties. Tech. Rep. 4, USAF School of Aviation Medicine, Randolph Field, Texas (1951)
7. Fukunaga, K., Hostetler, L.D.: k-nearest neighbor Bayes risk estimation. IEEE Transactions on Information Theory 21, 285–293 (1975)
8. Jiao, L., Denœux, T., Pan, Q.: Fusion of pairwise nearest-neighbor classifiers based on pairwise-weighted distance metric and Dempster-Shafer theory. In: Proceeding of 17th International Conference on Information Fusion. pp. 1–7 (2014)
9. Kanj, S., Abdallah, F., Denœux, T., Tout, K.: Editing training data for multi-label classification with the k-nearest neighbor rule. Pattern Analysis and Applications, published on line, DOI 10.1007/s10044–015–0452–8 (2015)
10. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy K-nearest neighbor algorithm. IEEE Transactions on Systems, Man and Cybernetics 15, 580–585 (1985)
11. Koplowitz, J., Brown, T.A.: On the relation of performance to editing in nearest neighbor rules. Pattern Recognition 13, 251–255 (1981)
12. Liu, Z., Pan, Q., Dezert, J.: A new belief-based k-nearest neighbor classification method. Pattern Recognition 46, 834–844 (2013)
13. Merz, C.J., Murphy, P.M., Aha, D.W.: UCI Repository of Machine Learning Databases. Department of Information and Computer Science, University of California, Irvine. http://www.ics.uci.edu/mlearn/MLRepository.html (1997)
14. Shafer, G.: A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press (1976)
15. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. International Journal of Approximate Reasoning 38, 133–147 (2005)
16. Wilson, D.L.: Asymptotic properties of nearest neighbor rules using edited data sets. IEEE Transactions on Systems, Man and Cybernetics 2, 408–421 (1972)
17. Zadeh, L.A.: Fuzzy sets. Information Control 8, 338–353 (1965)