A Fuzzy Preference-Based Dempster-Shafer Evidence Theory for Decision Fusion

Chaosheng Zhu[a,1], Bowen Qin[a,1], Fuyuan Xiao[a,*], Zehong Cao[b,1], Hari Mohan Pandey[c]

[a]School of Computer and Information Science, Southwest University, No.2 Tiansheng Road, BeiBei District, Chongqing 400715, China
[b]Discipline of ICT, School of Technology, Environments and Design, University of Tasmania, Hobart, TAS, Australia
[c] Department of Computer Science, Edge Hill University, UK

Abstract

Dempster-Shafer evidence theory (D-S) is an effective instrument for merging the collected pieces of basic probability assignment (BPA), and it exhibits superiority in achieving robustness of soft computing and decision making in an uncertain and imprecise environment. However, the determination of BPA is still uncertain, and merely applying evidence theory can sometimes lead to counterintuitive results when lines of evidence conflict. In this paper, a novel BPA generation method for binary problems called the base algorithm is designed based on the kernel density estimation to construct probability density function models, using the pairwise learning method to establish binary classification pairs. By means of the new BPA generation method, a new decision-making algorithm based on D-S evidence theory, fuzzy preference relation and nondominance criterion is effectively designed. The strength of

*Corresponding author: Fuyuan Xiao, School of Computer and Information Science, Southwest University, Chongqing 400715, China.

  Email addresses: `zhuchaoshengswu@163.com` (Chaosheng Zhu), `qinbowen_swu@163.com` (Bowen Qin), `doctorxiaofy@hotmail.com` (Fuyuan Xiao), `Zehong.Cao@uts.edu.au` (Zehong Cao), `pandeyh@edgehill.ac.uk` (Hari Mohan Pandey)

[1]These authors contributed equally to this paper.

the proposed method is in applying pairwise learning, which transforms the original complex problem into simple subproblems. With this process, the complexity of the problem to be solved is greatly reduced, which increases the feasibility for industrial fields. Furthermore, the fuzzy computing technique is used to aggregate the output for each single subproblem, and the nondominance degree of each class is determined from the fuzzy preference relation matrix, which can be directly used for the determination of the input instance. Based on several industrial-based classification experiments, the proposed BPA generation method and decision-making algorithm present the effectiveness and improvement in terms of precision and Cohen's kappa.

1. Introduction

Information fusion can integrate data from multiple sources to obtain the optimum estimation of targets [1, 2]. The method has been applied in many fields, including medical diagnosis [3, 4], risk analysis [5], uncertainty problems [6] and multicriteria decision-making [7–10]. The information in a multisource system may be imprecise, uncertain or even mutually conflicting in real applications. Numerous theories are presented to handle uncertainty [11, 12], including intuitionistic fuzzy sets [13], soft likelihood function [14], soft sets [15], evidence theory [16, 17] and others [18–20].

D-S evidence theory has been widely applied in numerous fields owing to its superior features [21]. Multisource information can be fused by Dempster's combination rule without depending on prior information [22]. D-S evidence theory can handle uncertain and unknown information with fewer conditions than probability theory.

Due to its flexibility, many other theories can be extended and combined with D-S evidence theory in its elegant computational framework, such as evidential reasoning, D numbers, Z-numbers [23–25], belief rule, quantum mass function, complex evidence theory [26], and other hybrid models [27, 28]. These combination methods provide more extensive application of D-S evidence theory and generate various valuable approaches to solve many problems, such as decision making [29–31], reliability evaluation [32], classification [33], information fusion and so on [34–36].

However, D-S evidence theory still presents some shortcomings [37]. As the number of elements in the frame of discernment increases, so does the computational complexity. In addition, when evidence is highly conflicting, the fusion result may be abnormal; Zadeh cites a famous example to argue this issue [38]. It can be noticed that these issues are all related to BPA, because the determination of BPA is the first step, which directly impacts the computation of D-S evidence theory. Therefore, this is troublesome for cases which must be solved urgently since the determination of BPA is not always reasonable. When combining these BPAs, the BPAs may create a negative impression and may ultimately influence the decision-making result. The effectiveness of the relevant evidential system directly depends on the correctness of the generated BPAs. This is therefore worth further study, especially for real-world applications.

A large number of methods have been presented for the purpose of addressing the problem mentioned above. Suh and Yook [39] used sensor data to determine BPA. Jiang et al. made use of the fuzzy Triangle number to construct the BPA [40]. Xu et al. [41] presented a nonparametric way to generate BPA. Qin et al. [42] improved BPA by introducing kernel density estimation. Nevertheless, these methods are not stable enough since they do not consider complex cases or cases with multiple targets. For instance, when addressing multiclass problems, most of them cannot clearly

distinguish the different boundaries of different classes. In multiclass problems, with the increase of the number of classes, the computational complexity will significantly increase since Dempster's combination rule includes exponential behavior. These approaches thus still have significant room for improvement. Therefore, a new BPA determination method for binary problems, the base algorithm of which is based on the pairwise learning strategy and kernel density estimation-based probability density function, is proposed in this paper. It can be used to determine BPA in a more reasonable and effective way.

In particular, the main point is transforming the original individual problem into binary problems. Correspondingly, the frame of discernment is transformed into multiple binary hypotheses. These binary hypotheses can then be addressed separately. Compared with dedicated multiclass classifiers, binary classifiers are generally easier to construct, train and test more rapidly, and they have simpler decision boundaries. To solve the complex problems that often exist in computer vision with plenty of targets or high feature dimensions, a more practical and feasible choice is usually to divide multiple classes of problems into multiple binary classification problems that are easily solved. In addition, in each sub-binary problem, the kernel density estimation is applied to construct the probability density function (PDF) models. The BPAs of each binary problem can be obtained according to the relation between test samples and PDF models. After modeling the fuzzy preference relationship for the obtained BPAs, nondominance degree is computed by comparing the nondominance degree of each class, and the class to which the instance belongs is determined. Several classification experiments are implemented to demonstrate the effectiveness of the presented study. To evaluate the results, both Cohen's kappa and classification accuracy are under consideration.

The remaining parts of the study are presented as follows. In Section 2, the pre-

4

liminaries of this paper are quickly introduced with respect to D-S evidence theory, pairwise learning and kernel density estimation. The details of the proposed base algorithm for binary problems based on the pairwise learning algorithm and kernel density estimation-based PDF model are presented, and the novel decision-making method based on the D-S evidence theory, fuzzy preference relation and nondominance criterion is proposed in Section 3. Section 4 shows an application of the proposed method, and a few classification experiments are introduced in Section 5. Finally, the conclusion to this work is presented in Section 6.

2. Preliminaries

In this section, the preliminaries of this paper are shortly introduced with respect to D-S evidence theory, pairwise learning and kernel density estimation.

2.1. Dempster-Shafer evidence theory

The Dempster-Shafer evidence theory [43, 44] is one of the most effective methods. With the promotion of Bayesian probability theory, D-S evidence theory offers the superior capability that it can directly express the "uncertainty" through assigning belief into multi-event hypotheses rather than into single-event hypotheses alone. The basic notions about the D-S evidence theory are presented in the following.

2.1.1. Frame of discernment

Let $\mathcal{C}$ be a set of k collectively exhaustive and mutually exclusive events (denoted as C), where $\mathcal{C}$ is known as the frame of discernment [43, 44]:

$$\mathcal{C} = \left\{ C_1, C_2, \cdots, C_p, \cdots C_q, \cdots C_k \right\}. \tag{1}$$

$2^{\mathcal{C}}$ denotes the possible hypotheses formed by the combination of classes in the $\mathcal{C}$:

$$2^{\mathcal{C}} = \{\emptyset, \{C_1\}, \{C_2\}, \cdots, \{C_k\}, \{C_1, C_2\}, \cdots, \{C_p, C_q\}, \cdots, \mathcal{C}\}, \qquad (2)$$

where $\emptyset$ denotes the empty set. The hypothesis which contains only one class is called singleton.

2.1.2. Basic probability assignment

Once $\mathcal{C}$ is established, the BPA m, also called the mass function, is defined in [43, 44] as:

$$m : \begin{cases} 2^{\mathcal{C}} & \to [0,1], \\ A & \to m(A), \end{cases} \qquad (3)$$

and:

$$\sum_{A \in 2^{\mathcal{C}}} m(A) = 1, \qquad (4)$$

$$m(\emptyset) = 0, \qquad (5)$$

where $m(A)$ is the support degree to hypothesis A. A is called a focal element if $m(A) > 0$.

Owing to its superior feature of modeling uncertainty, BPA has been deeply explored with respect to entropy [45], negation, divergence, and information quality.

6

2.1.3. Dempster's combination rule

Let m_1 and m_2 be two mass functions separately collected from distinctive information sources on $\mathcal{C}$. The combined evidence $m = m_1 \oplus m_2$ is calculated as follows:

$$m(A) = \frac{\sum_{B \cap D = A} m_1(B) m_2(D)}{1 - \mathbb{K}}, \tag{6}$$

and

$$\mathbb{K} = \sum_{B \cap D = \emptyset} m_1(B) m_2(D), \tag{7}$$

where $B, D \in 2^{\mathcal{C}}$; $\mathbb{K}$ is the conflict coefficient between B and D [46].

2.1.4. Pignistic probability transformation

Pignistic probability is an elaboration of the D-S evidence theory developed by Smets. The pignistic probability employs the principle of insufficient reason to transform the mass function into a probability measure for decision making, which provides a valid way to address the nonprobability problem in D-S evidence theory.

Let m be a BPA defined on $\mathcal{C}$, where the pignistic probability function of hypothesis B in the frame $\mathcal{C}$ is defined in [47] as:

$$BetP(B) = \sum_{\substack{A \subseteq 2^{\mathcal{C}} \\ A \cap B \neq \emptyset}} \frac{|A \cap B|}{|A|} \cdot \frac{m(A)}{1 - m(\emptyset)}, \quad \forall B \subseteq 2^{\mathcal{C}}, \tag{8}$$

where $|A|$ is the cardinality of hypothesis A. $BetP(B)$ describes all the probability values that support the hypothesis B being true.

2.2. Pairwise learning

The pairwise learning method is a useful method to handle multiclass classification problems through class binarization techniques. The main strategy of pairwise

learning is to transform the original multiclass classification problem into multiple binary classification problems, which are easier to distinguish [48]. In the pairwise learning approach, the original k-classes classification problem is divided into $\frac{k(k-1)}{2}$ binary classification problems. Then, a binary classifier is trained to distinguish each pair of binary classification problems using a predefined binary classification algorithm. In this way, $\frac{k(k-1)}{2}$ binary classifiers can be obtained based on the training dataset, where each classifier consists of only two class labels. An example of pairwise learning for a multiclass classification problem is presented in Fig. 1.

For a test sample to be classified, the final classification result is determined by combining the outputs of all binary classifiers. A score-matrix R containing these outputs is constructed as:

$$
R = \begin{bmatrix}
- & \cdots & r_{(1,p)} & \cdots & r_{(1,q)} & \cdots & r_{(1,k)} \\
\cdots & - & \cdots & \cdots & \cdots & \cdots & \cdots \\
r_{(p,1)} & \cdots & - & \cdots & r_{(p,q)} & \cdots & r_{(p,k)} \\
\cdots & \cdots & \cdots & - & \cdots & \cdots & \cdots \\
r_{(q,1)} & \cdots & r_{(q,p)} & \cdots & - & \cdots & r_{(q,k)} \\
\cdots & \cdots & \cdots & \cdots & \cdots & - & \cdots \\
r_{(k,1)} & \cdots & r_{(k,p)} & \cdots & r_{(k,q)} & \cdots & -
\end{bmatrix},
\tag{9}
$$

where $r_{(p,q)}$ is the confidence that the binary classifier distinguishes class C_p from class C_q, and $r_{(q,p)} = 1 - r_{(p,q)}$ is the confidence in favor of class C_q (if it is not provided by the classifier). After the score-matrix is determined by Eq. (9), the next step is to deduce the final class by any of the combination methods. In this paper, the nondominance criterion (ND) based on fuzzy preference relations [48] is applied. The score-matrix is regarded as a fuzzy preference relation, and any

degree of nondominance is calculated from the fuzzy preference relation. Specifically, nondominance represents the degree to which the class C_p is not dominated by any of the remaining classes. The detailed classification steps are as follows:

1. Normalize the score-matrix R, which is denoted as R':

$$
R' = \begin{bmatrix}
- & \cdots & r'_{(1,p)} & \cdots & r'_{(1,q)} & \cdots & r'_{(1,k)} \\
\cdots & - & \cdots & \cdots & \cdots & \cdots & \cdots \\
r'_{(p,1)} & \cdots & - & \cdots & r'_{(p,q)} & \cdots & r'_{(p,k)} \\
\cdots & \cdots & \cdots & - & \cdots & \cdots & \cdots \\
r'_{(q,1)} & \cdots & r'_{(q,p)} & \cdots & - & \cdots & r'_{(q,k)} \\
\cdots & \cdots & \cdots & \cdots & \cdots & - & \cdots \\
r'_{(k,1)} & \cdots & r'_{(k,p)} & \cdots & r'_{(k,q)} & \cdots & -
\end{bmatrix}, \tag{10}
$$

in which

$$
r'_{(p,q)} = \frac{r_{(p,q)}}{r_{(p,q)} + r_{(q,p)}}. \tag{11}
$$

2. Construct the fuzzy preference relation matrix $\widetilde{R}$:

$$
\widetilde{R} = \begin{bmatrix}
- & \cdots & \widetilde{r}_{(1,p)} & \cdots & \widetilde{r}_{(1,q)} & \cdots & \widetilde{r}_{(1,k)} \\
\cdots & - & \cdots & \cdots & \cdots & \cdots & \cdots \\
\widetilde{r}_{(p,1)} & \cdots & - & \cdots & \widetilde{r}_{(p,q)} & \cdots & \widetilde{r}_{(p,k)} \\
\cdots & \cdots & \cdots & - & \cdots & \cdots & \cdots \\
\widetilde{r}_{(q,1)} & \cdots & \widetilde{r}_{(q,p)} & \cdots & - & \cdots & \widetilde{r}_{(q,k)} \\
\cdots & \cdots & \cdots & \cdots & \cdots & - & \cdots \\
\widetilde{r}_{(k,1)} & \cdots & \widetilde{r}_{(k,p)} & \cdots & \widetilde{r}_{(k,q)} & \cdots & -
\end{bmatrix}, \tag{12}
$$

where the fuzzy preference relation $\widetilde{r}_{(p,q)}$ is calculated as below:

$$\widetilde{r}_{(p,q)} = \begin{cases} r'_{(p,q)} - r'_{(q,p)}, & r'_{(p,q)} > r'_{(q,p)}; \\ 0, & otherwise. \end{cases}$$

3. Compute the nondominance degree ND_p of each class C_p:

$$ND_p = 1 - \sup_{1 \le p \le k} [\widetilde{r}_{(q,p)}]. \tag{13}$$

4. Determine the final class:

Select the max value in ND_p, denoted as $\mathcal{ND}$, which indicates that the instance belongs to class $C_{\mathcal{ND}}$:

$$\mathcal{ND} = \arg \max_{1 \le p \le k} \{ND_p\}. \tag{14}$$

Hence, the class of the input instance is finally identified as belonging to the class $C_{\mathcal{ND}}$.

2.3. Kernel density estimation

Nonparametric density estimation is a considerable instrument for statistical analysis of data. The nonparametric approach exhibits the superior capability that it can more flexibly model the given dataset. Furthermore, compared with traditional approaches, the nonparametric approach will not be affected by specification bias.

Kernel density estimation (KDE), as a nonparametric density estimation approach, is used to estimate the probability density function (PDF) of a random variable. For a univariate independent training sample $T = (t_1, t_2, t_3 \cdots, t_n)$, the KDE
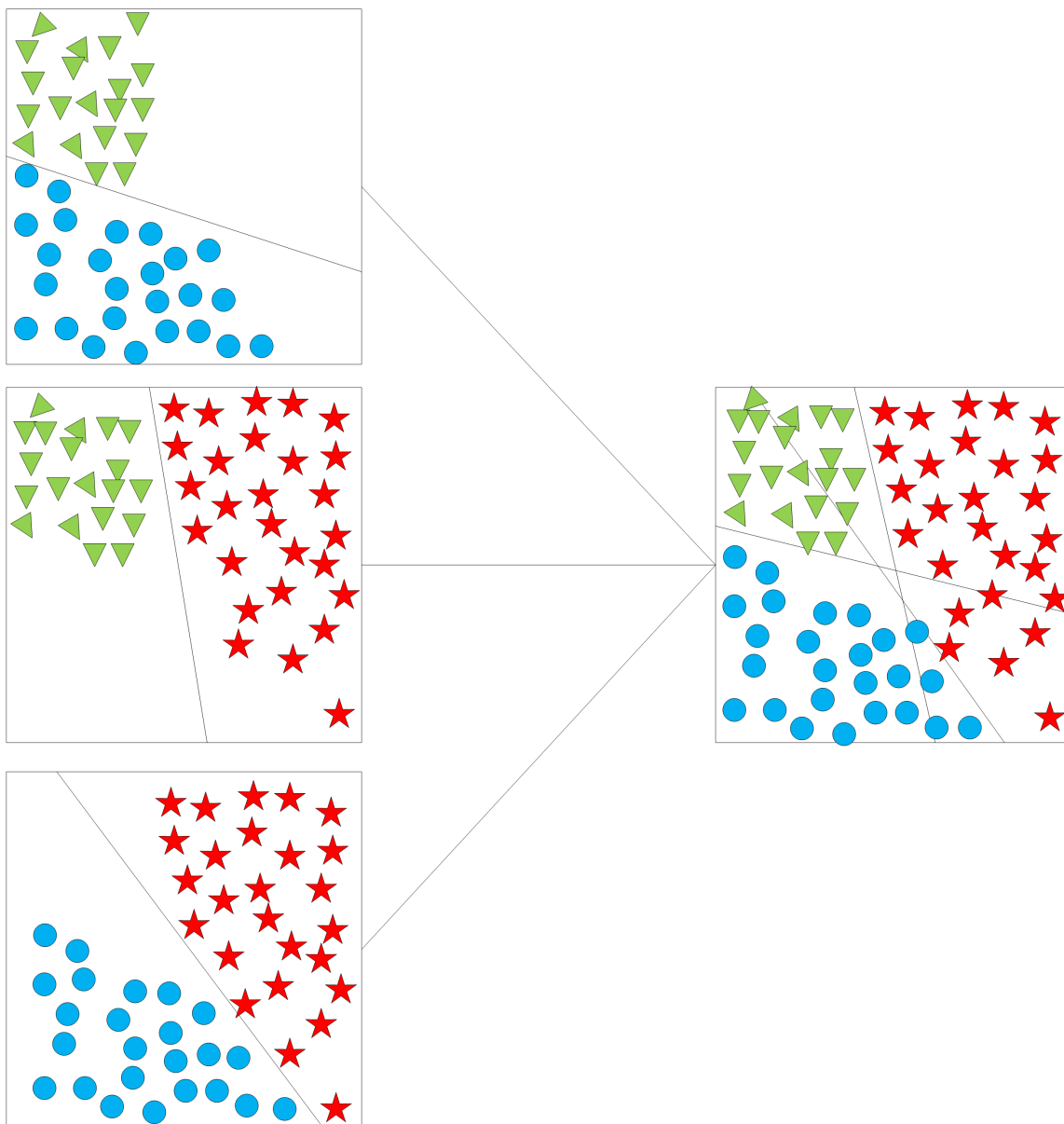
Figure 1: An example of the pairwise learning technique for a 3-class problem.

of the density which is implemented to construct the probability density function (PDF) of test sample T is defined as:

$$\overset{\wedge}{f_h}(x) = \frac{1}{n} \sum_{z=1}^{n} \kappa_h(x - t_z) = \frac{1}{nh} \sum_{z=1}^{n} \kappa(\frac{x - t_z}{h}), \quad x \in \mathbb{R}, \tag{15}$$

where κ is the kernel function; h is a smoothing parameter called the bandwidth; $\mathbb{R}$ is any real number.

In this study, an adaptive KDE approach [42] is chosen, and its corresponding knowledge is as follows:

The KDE can be computed by:

$$\hat{f}(x; h) = \frac{1}{n} \sum_{z=1}^{n} \kappa(x, t_z; h), \quad x \in [0, 1], \tag{16}$$

where the kernel κ is:

$$\kappa(x, t_z; h) = \sum_{K=-\infty}^{\infty} \phi(x, 2K + t_z; h) + \phi(x, 2K - t_z; h), \tag{17}$$

and

$$\phi(x, 2K + t_z; h) = \frac{1}{\sqrt{2\pi h}} e^{\frac{-(x-2K-t_z)^2}{2h}}, \tag{18}$$

$$\phi(x, 2K - t_z; h) = \frac{1}{\sqrt{2\pi h}} e^{\frac{-(x-2K+t_z)^2}{2h}}. \tag{19}$$

For more details of KDE, the reader can refer to [42].

3. The proposed method

In this section, a new method for the determination of BPA is presented. Furthermore, a new decision-making algorithm based on D-S evidence theory is proposed.

In general, in pairwise learning, the original frame of discernment which contains k classes is divided into $\dfrac{k(k-1)}{2}$ hypotheses pairs. For instance, given a problem with the frame of discernment $\Theta = \{A, B, D\}$, it can be decomposed to the hypothesis pairs of $\{A, B\}$, $\{A, D\}$ and $\{B, D\}$, which constitute three smaller binary problems. It can be determined that a complex problem with multiple classes included is decomposed into multiple easier binary problems, for which each frame of discernment of binary problems includes only two classes/events. The newly devised base algorithm can then be applied to each binary problem to determine BPA, and the BPA at this time is the immediate result for binary problems. Subsequently, their results will be aggregated by the fuzzy preference relation to determine the final BPA. Note that the obtained BPA is for the original problem.

3.1. Decompose the frame of discernment

For the given dataset, if it contains k classes and each class has N attributes, then the frame of discernment of the dataset is defined:

$$\mathcal{C} = \{C_1, C_2, \cdots, C_p, \cdots C_q, \cdots C_k\}, \tag{20}$$

which means that there are k classes in the dataset. Conducting the decomposition process will generate $\frac{k(k-1)}{2}$ hypothesis pairs, denoted as θ:

$$\theta_1 = \{C_1, C_2\}, \theta_2 = \{C_1, C_3\}, \ldots, \theta_i = \{C_p, C_q\}, \ldots \theta_{\frac{k(k-1)}{2}} = \{C_{k-1}, C_k\}. \tag{21}$$

After that, the original k-classes problem is divided into $\frac{k(k-1)}{2}$ binary problems, and the frames of discernment are all converted to a binary problem hypothesis which contains two classes. For each binary problem, the base algorithm introduced below is applied to determine the BPA.

13

3.2. Base algorithm

In this process, the base algorithm for each binary problem is divided into three parts. The first part builds a probability density function (PDF) model via kernel density estimation introduced in Section 2.3. The second part determines the BPA with the relation between the test instance and the model. The final part makes a final decision based on the fuzzy preference relation and the nondominance degree.

3.2.1. Build the probability density function (PDF) model

Step 1: Preliminary preparation and hypothesis

First, the whole dataset is divided into two parts: training set $T = \{t_1, t_2, \ldots, t_n\}$ and test set $S = \{s_1, s_2, \ldots, s_v\}$. The class to which instance t_n/s_v belongs is one of the classes in $\mathcal{C}$, and for each instance t_n/s_v, N attributes are included.

Step 2: Construct the PDF model

The lowest value LB_{pj} and highest value UB_{pj} of the training sample $t_{p,j}$ in training set T for class C_p and its corresponding attribute j are selected to construct the interval $[LB_{pj}, UB_{pj}]$.

If $t_z \in [LB_{pj}, UB_{pj}]$:

$$\hat{f}(x; h) = \frac{1}{n} \sum_{z=1}^{n} \kappa(x, t_z; h), \quad x \in [0, 1]; \tag{22}$$

If $t_z \notin [LB_{pj}, UB_{pj}]$:

$$\hat{f}(x; h) = \varepsilon_0, \tag{23}$$

where, according to paper [42], ε_0 is set as 0.001.

The PDF model of the given dataset has already been established.

3.2.2. Determine BPA for the binary problem $\{C_p, C_q\}$

For the binary problem $\{C_p, C_q\}$, $(1 \leq p < q \leq k)$, the determination of its BPA is as follows.

Step 3: Obtain the relationship between test sample s_v and the PDF model

Use y_l $(l = 1, 2)$ to represent the intersection of the selected attribute j $(j = 1, 2, \cdots, N)$ with the corresponding PDF models of two classes in the hypothesis pair $\{C_p, C_q\}$, $(1 \leq p < q \leq k)$.

Step 4: Assign the membership to the focal element

Case 1: $\exists y_l \neq \varepsilon_0$ $(l = 1, 2)$

The intersecting values y_l $(l = 1, 2)$ are sorted in descending order, w_g $(g = 1, 2)$. For w_g $(g = 1, 2)$, its corresponding class (i.e., the class of PDF model it belongs to) can be identified in the binary problem $\{C_p, C_q\}$. For the hypothesis pair, w_g $(g = 1, 2)$ is assigned to the mass function by the following rule:

$$m_{(p,q),j}(\{C_p\}) = w_1,$$

$$m_{(p,q),j}(\{C_p, C_q\}) = w_2,$$

where $m_{(p,q),j}(\{C_p\})$ denotes the w_g assignment of the mass function to the hypothesis pair for class C_p to attribute j in the hypothesis pair $\{C_p, C_q\}$.

Case 2: $\forall y_l = \varepsilon_0$ $(l = 1, 2)$

In this case, we generate the BPA by considering the similarity between the test sample s_v and the mean value of each class in training set T. Let $\overline{t_{p,j}}$ denote the mean value of class C_p for attribute j in T, $\overline{t_{p,j}} = \dfrac{\sum\limits_{p=1}^{k} t_{p,j}}{k}$, in which $t_{p,j}$ denotes the sample of class C_p for attribute j, and k is the cardinality of the frame of discernment $\mathcal{C}$. With

15

the input instance $s_{v,j}$, which is the instance s_v with its corresponding attribute j, the distance $d_{p,j}$ from the instance to the $\overline{t_{p,j}}$ is calculated as follows, which indicates the discrepancy of the instance with respect to all classes.

For class C_p $(1 \leq p \leq k)$:

$$d_{p,j} = \left| s_{v,j} - \overline{t_{p,j}} \right|, \tag{24}$$

where j represents the attribute j, $1 \leq j \leq N$.

If $max(d_{p,j}) > 1$, $d_{p,j}{}'$ is calculated as follows:

$$d_{p,j}{}' = \frac{d_{p,j}}{d_{p,j} + d_{q,j}}. \tag{25}$$

If $max(d_{p,j}) < 1$, $d_{p,j}{}'$ is calculated as follows:

$$d_{p,j}{}' = d_{p,j}. \tag{26}$$

Then, $m_{(p,q),j}(\{C_p\})$ can be obtained by:

$$m_{(p,q),j}(\{C_p\}) = e^{-d_{p,j}{}'}. \tag{27}$$

Similarly, for the class C_q and its corresponding attribute j, the same method is conducted as above to obtain $m_{(p,q),j}(\{C_q\})$.

Then, the BPAs of the test sample can be determined.

Step 5: Normalize the BPA

Due to the requirements of Eq. (6), the sum of the belief of all hypotheses should be equal to 1; thus, the BPAs should be normalized by:

Case 1: $\exists y_l \neq \varepsilon_0 \ (l = 1, 2)$

$$m'_{(p,q),j}(\{C_p\}) = \frac{m_{(p,q),j}(\{C_p\})}{m_{(p,q),j}(\{C_p\}) + m_{(p,q),j}(\{C_p, C_q\})}. \tag{28}$$

$$m'_{(p,q),j}(\{C_p, C_q\}) = \frac{m_{(p,q),j}(\{C_p, C_q\})}{m_{(p,q),j}(\{C_p\}) + m_{(p,q),j}(\{C_p, C_q\})}. \tag{29}$$

Case 2: $\forall y_l = \varepsilon_0 \ (l = 1, 2)$

$$m'_{(p,q),j}(\{C_p\}) = \frac{m_{(p,q),j}(\{C_p\})}{m_{(p,q),j}(\{C_p\}) + m_{(p,q),j}(\{C_q\})}. \tag{30}$$

$$m'_{(p,q),j}(\{C_q\}) = \frac{m_{(p,q),j}(\{C_q\})}{m_{(p,q),j}(\{C_p\}) + m_{(p,q),j}(\{C_q\})}. \tag{31}$$

Note that the generated BPA is not the final one, but is for each binary problem.

Step 6: Combine BPAs

These BPAs can be fused by Dempster's combination rule for each binary problem according to Eq. (6) and Eq. (7) to obtain the final BPA ($\widetilde{m}$) of the sample.

$$\widetilde{m}_{(p,q)} = m'_{(p,q),1} \oplus m'_{(p,q),2} \oplus, \cdots, \oplus m'_{(p,q),j}, \cdots, \oplus m'_{(p,q),N}. \tag{32}$$

Then, for each binary problem $\{C_p, C_q\}$, its corresponding BPA can be obtained as $\widetilde{m}_{(p,q)}$.

With the obtained final BPA $\widetilde{m}_{(p,q)}$, the pignistic probability function is implemented as defined in Eq. (8) to transform the mass function to a probability distribution:

$$BetP_{(p,q)}(\{C_p\}) = \widetilde{m}_{(p,q)}(\{C_p\}) + \frac{|\{C_p\} \cap \{C_p, C_q\}|}{|\{C_p, C_q\}|} \cdot \frac{\widetilde{m}_{(p,q)}(\{C_p, C_q\})}{1 - \widetilde{m}_{(p,q)}(\varnothing)}; \tag{33}$$

17

$$BetP_{(p,q)}(\{C_q\}) = \frac{|\{C_q\} \cap \{C_p, C_q\}|}{|\{C_p, C_q\}|} \cdot \frac{\widetilde{m}_{(p,q)}(\{C_p, C_q\})}{1 - \widetilde{m}_{(p,q)}(\varnothing)}. \tag{34}$$

It is noteworthy that the problem to be addressed occurs in a closed world, which indicates that $\mathcal{C}$ is complete; thus, $\widetilde{m}(\varnothing) = 0$.

3.2.3. Aggregation and final decision making

Step 7: Determine the score-matrix

As mentioned above, $r_{(p,q)}$ represents the confidence degree of the classifier in favor of class C_p. In the proposed method, $r_{(p,q)}\{C_p\}$ is obtained by $r_{(p,q)}\{C_p\} = BetP_{(p,q)}(\{C_p\})$, and $r_{(q,p)}\{C_q\}$ is assigned by $r_{(q,p)}\{C_q\} = BetP_{(q,p)}(\{C_q\})$, where $BetP_{(p,q)}(\{C_p\})$ and $BetP_{(q,p)}(\{C_q\})$ are the instances of $BetP$ supporting class C_p and class C_q, respectively. Thus, in this paper, the score-matrix can be obtained as:

$$R = \begin{bmatrix} - & \cdots & r_{(1,p)}\{C_1\} & \cdots & r_{(1,q)}\{C_1\} & \cdots & r_{(1,k)}\{C_1\} \\ \cdots & - & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{(p,1)}\{C_p\} & \cdots & - & \cdots & r_{(p,q)}\{C_p\} & \cdots & r_{(p,k)}\{C_p\} \\ \cdots & \cdots & \cdots & - & \cdots & \cdots & \cdots \\ r_{(q,1)}\{C_q\} & \cdots & r_{(q,p)}\{C_q\} & \cdots & - & \cdots & r_{(q,k)}\{C_q\} \\ \cdots & \cdots & \cdots & \cdots & \cdots & - & \cdots \\ r_{(k,1)}\{C_k\} & \cdots & r_{(k,p)}\{C_k\} & \cdots & r_{(k,q)}\{C_k\} & \cdots & - \end{bmatrix}, \tag{35}$$

where $r_{(p,q)}\{C_p\}$ represents the pignistic probability supporting the class C_p in the frame of discernment $\{C_p, C_q\}$.

Step 8: Normalize the score-matrix

$$
R' = \begin{bmatrix}
- & \cdots & r'_{(1,p)}\{C_1\} & \cdots & r'_{(1,q)}\{C_1\} & \cdots & r'_{(1,k)}\{C_1\} \\
\cdots & - & \cdots & \cdots & \cdots & \cdots & \cdots \\
r'_{(p,1)}\{C_p\} & \cdots & - & \cdots & r'_{(p,q)}\{C_p\} & \cdots & r'_{(p,k)}\{C_p\} \\
\cdots & \cdots & \cdots & - & \cdots & \cdots & \cdots \\
r'_{(q,1)}\{C_q\} & \cdots & r'_{(q,p)}\{C_q\} & \cdots & - & \cdots & r'_{(q,k)}\{C_q\} \\
\cdots & \cdots & \cdots & \cdots & \cdots & - & \cdots \\
r'_{(k,1)}\{C_k\} & \cdots & r'_{(k,p)}\{C_k\} & \cdots & r'_{(k,q)}\{C_k\} & \cdots & -
\end{bmatrix}, \quad (36)
$$

in which

$$
r'_{(p,q)}\{C_p\} = \frac{r_{(p,q)}\{C_p\}}{r_{(p,q)}\{C_p\} + r_{(q,p)}\{C_q\}}. \quad (37)
$$

Step 9: Compute the fuzzy preference relation

$$
\widetilde{R} = \begin{bmatrix}
- & \cdots & \widetilde{r}_{(1,p)}\{C_1\} & \cdots & \widetilde{r}_{(1,q)}\{C_1\} & \cdots & \widetilde{r}_{(1,k)}\{C_1\} \\
\cdots & - & \cdots & \cdots & \cdots & \cdots & \cdots \\
\widetilde{r}_{(p,1)}\{C_p\} & \cdots & - & \cdots & \widetilde{r}_{(p,q)}\{C_p\} & \cdots & \widetilde{r}_{(p,k)}\{C_p\} \\
\cdots & \cdots & \cdots & - & \cdots & \cdots & \cdots \\
\widetilde{r}_{(q,1)}\{C_q\} & \cdots & \widetilde{r}_{(q,p)}\{C_q\} & \cdots & - & \cdots & \widetilde{r}_{(q,k)}\{C_q\} \\
\cdots & \cdots & \cdots & \cdots & \cdots & - & \cdots \\
\widetilde{r}_{(k,1)}\{C_k\} & \cdots & \widetilde{r}_{(k,p)}\{C_k\} & \cdots & \widetilde{r}_{(k,q)}\{C_k\} & \cdots & -
\end{bmatrix}, \quad (38)
$$

where the fuzzy preference relation $\widetilde{r}_{(p,q)}$ is calculated as below:

$$\widetilde{r}_{(p,q)}\{C_p\} = \begin{cases} r'_{(p,q)}\{C_p\} - r'_{(q,p)}\{C_q\}, & r'_{(p,q)}\{C_p\} > r'_{(q,p)}\{C_q\}; \\ 0, & otherwise. \end{cases} \tag{39}$$

Step 10: Compute the nondominance degree ND_p of class C_p

$$ND_p = \begin{cases} 1 - \sup_{1 \le q \le k} [\widetilde{r}_{(q,p)}\{C_q\}], & 1 - \sup_{1 \le q \le k} [\widetilde{r}_{(q,p)}\{C_q\}] > 0; \\ 0, & 1 - \sup_{1 \le q \le k} [\widetilde{r}_{(q,p)}\{C_q\}] < 0. \end{cases} \tag{40}$$

Step 11: Determine the final class of the instance

Select the maximum value in ND_p, denoted as $\mathcal{ND}$, which indicates that the instance belongs to class $C_{\mathcal{ND}}$:

$$\mathcal{ND} = \arg\max_{1 \le p \le k}\{ND_p\}. \tag{41}$$

Hence, the input instance is finally identified as belonging to the class $C_{\mathcal{ND}}$.

4. An illustrated application

Decision making is necessary in various kinds of applications [49]. In this section, an application of classification is given to illustrate the proposed method. The Iris dataset originates from the UCI repository and KEEL dataset repository of the machine learning databases. The Iris dataset includes three species of Setosa (Se), Versicolor (Vc), and Virginica (Vi), with 50 instances per class. Each type of flower has four attributes: sepal length (SL), sepal width (SW), petal length (PL), and petal width (PW).
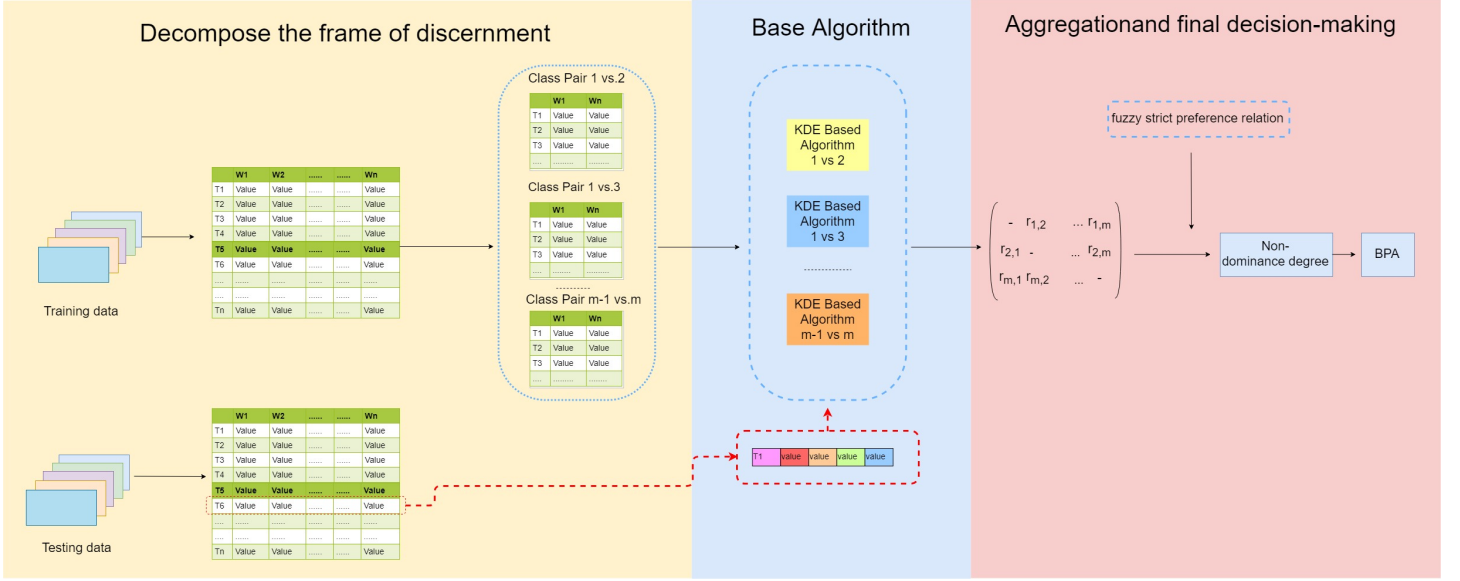
Figure 2: The flowchart of the proposed method

Step 1

For the Iris dataset, there are 30 and 20 instances of each class for the training set and test set, respectively.

Step 2

For each attribute of each class, the probability density function (PDF) modes will be constructed based on KDE. The PDF model of the property SL of the class Setosa is established in Fig. 3(a). Similarly, for the attributes SW, PL and PW of Setosa, their corresponding PDF is established and drawn in Fig. 3(b)-(d), respectively. In addition, for the classes Versicolor and Virginicia, their PDF probability model can be established with the same operation.

Step 3

In this step, we demonstrate calculating the confidence degree of $r_{(Se,Vc)}$. Given an instance of class Setosa from the test set: $s_{v,SL} = 5.1$ cm, $s_{v,SW} = 3.8$ cm, $s_{v,PL}$
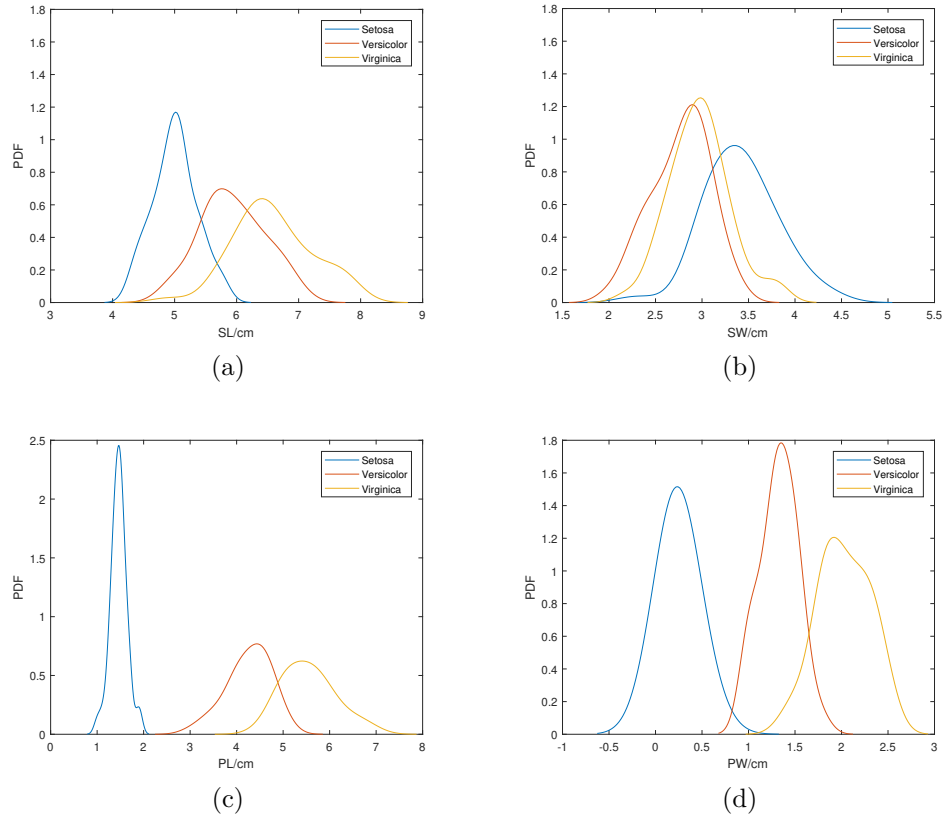
Figure 3: Probabilistic model of four attributes in three categories

$= 1.6$ cm, $s_{v,PW} = 0.2$ cm

To calculate its intersection with other PDF curves, consider the value of its SL attribute as $s_{v,SL} = 5.1$. The result is shown in Fig. 4. We can obtain two intersections A_1 and A_2 with the PDF curves of Setosa and Versicolor:

$$A_1 : (5.1, y_1) \text{ and } A_2 : (5.1, y_2).$$

The intersecting result of the test sample $s_{v,SL}$ with the PDF model is:

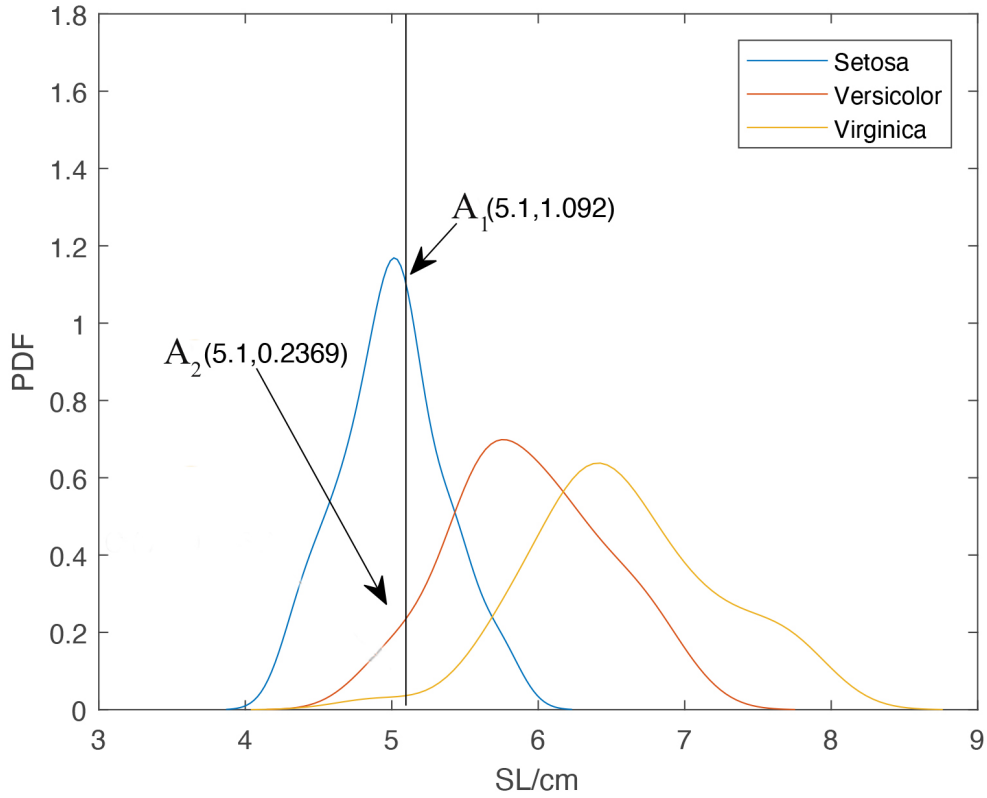$$y_1 = 1.0972 \text{ and } y_2 = 0.2369.$$

Figure 4: The intersection result with other PDF curves.

Step 4

Based on the rules in Section 3.2.2, we can obtain two propositions for the hypothesis pair $\{Se, Vc\}$:

$$m_{(Se,Vc)}(\{Se\}) = 1.0972;$$

$$m_{(Se,Vc)}(\{Se, Vc\}) = 0.2369.$$

Step 5

The BPA is normalized as:

$$m'_{(Se,Vc)}(\{Se\}) = 0.8224;$$

$$m'_{(Se,Vc)}(\{Se, Vc\}) = 0.1776.$$

Similarly, we obtain the remaining BPAs of attributes SW, PL and PW of hypothesis pair $\{Se, Vc\}$ with the proposed method, which are shown in Table 1. For the hypothesis pairs of $\{Se, Vi\}$ and $\{Vc, Vi\}$, their BPAs can be obtained with the same process as that employed for hypothesis pair $\{Se, Vc\}$.

Step 6

In this step, the four BPAs are merged using Dempster's merge rules. After that, the pignistic probability function is conducted to transform the BPA into probability distribution $BetP$. The results are $BetP_{(Se,Vc)}(\{Se\}) = 1$ and $BetP_{(Se,Vc)}(\{Vc\}) = 0$.

Step 7

To determine the score-matrix R, the same process above will be performed until all the items of R are calculated.

Table 1: The generated BPAs in terms of the test sample Setosa

Hypothesis	BPA
SL	$m'_{(Se,Vc)}(\{Se\}) = 0.8224$
	$m'_{(Se,Vc)}(\{Se,Vc\}) = 0.1776$
SW	$m'_{(Se,Vc)}(\{Se\}) = 0.9986$
	$m'_{(Se,Vc)}(\{Se,Vc\}) = 0.0014$
PL	$m'_{(Se,Vc)}(\{Se\}) = 0.9995$
	$m'_{(Se,Vc)}(\{Se,Vc\}) = 0.0005$
PW	$m'_{(Se,Vc)}(\{Se\}) = 0.9999$
	$m'_{(Se,Vc)}(\{Se,Vc\}) = 0.0001$

$$R = \begin{bmatrix} 0 & 1.000 & 1.0000 \\ 0 & 0 & 0.8278 \\ 0 & 0.1722 & 0 \end{bmatrix}. \tag{42}$$

Step 8

Normalize the score-matrix R to R':

$$R' = \begin{bmatrix} 0 & 1.000 & 1.0000 \\ 0 & 0 & 0.8278 \\ 0 & 0.1722 & 0 \end{bmatrix}. \tag{43}$$

Step 9

Compute the fuzzy preference relation between classes to obtain $\widetilde{R}$:

$$\widetilde{R} = \begin{bmatrix} 0 & 1.000 & 1.0000 \\ 0 & 0 & 0.6556 \\ 0 & 0 & 0 \end{bmatrix}. \tag{44}$$

25

Table 2: Datasets

d	Dataset	v	N	k
Hep	Hepatitis	155	19	2
Hea	Heart	270	13	2
Breast	Breast-cancer-wisconsin	699	9	2
Iri	Iris	150	4	3
Win	Wine	178	13	3
New	Newthyroid	215	5	3
Veh	Vehicle	846	18	4
Cle	Cleveland	303	13	5
Pag	Page-blocks	5472	10	5
Gla	Glass	214	9	7

Step 10

Compute the nondominance degrees:

$$ND_{Se} = 1;$$
$$ND_{Vc} = 0; \tag{45}$$
$$ND_{Vi} = 0.$$

Step 11

Select the maximum value in $\{ND_{Se}, ND_{Vc}, ND_{Vi}\}$, denoted as $\mathcal{ND}$, which indicates that the instance belongs to class $C_{\mathcal{ND}}$:

$$\mathcal{ND} = \arg \max_{p \in \{Se, Vc, Vi\}} \{ND_{Se}, ND_{Vc}, ND_{Vi}\} = Se. \tag{46}$$

Through $\mathcal{ND}$, it can then be inferred that the instance belongs to the class *Setosa*, which is the true discrimination of the class of the sample.

Table 3: Confusion matrix for an m-class problem

Ground label	Output class				
	C_1	C_2	$\cdots$	C_k	Total
C_1	μ_{11}	μ_{12}	$\cdots$	μ_{1k}	T_{r1}
C_2	μ_{21}	μ_{22}	$\cdots$	μ_{2k}	T_{r2}
...	$\cdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$
C_k	μ_{k1}	μ_{k2}	$\cdots$	μ_{kk}	T_{rk}
Total	T_{c1}	T_{c2}	$\cdots$	T_{ck}	T

Table 4: Confusion matrix of Heart dataset produced by the method of Jiang et al.

Actual class	Output class	
	1	2
1	135	15
2	52	68

Table 5: Confusion matrix of the Heart dataset produced by the method of Qin et al.

Actual class	Output class	
	1	2
1	146	4
2	47	73

Table 6: Average classification accuracy results of different methods for different datasets

Dataset	Jiang et al.	Qin et al.	Proposed method
Hep	81.29	89.03	91.61
Hea	75.19	81.11	91.48
Breast	86.38	97.07	99.85
Iri	95.33	96.00	99.33
Win	86.38	98.88	99.90
New	97.67	98.60	99.79
Veh	64.89	66.55	98.45
Cle	46.80	64.31	63.64
Pag	78.83	90.24	99.18
Gla	78.05	83.18	89.72

Table 7: Average kappa results of different datasets obtained with different methods

Dataset	Jiang et al.	Qin et al.	Proposed method
Hep	0.3787	0.6272	0.7632
Hea	0.4815	0.6003	0.8250
Breast	0.6733	0.9522	0.9968
Iri	0.9400	0.9500	0.9800
Win	0.9745	0.9830	0.9900
New	0.9600	0.9700	0.9800
Veh	0.4839	0.5208	0.9788
Cle	0.3643	0.4116	0.4731
Pag	0.5871	0.6181	0.9565
Gla	0.7425	0.7686	0.8574

Table 8: Confusion matrix of Heart dataset produced by the proposed method

Actual class	Output class	
	1	2
1	147	3
2	20	100

Table 9: Confusion matrix of the Page-blocks dataset produced by the method of Qin et al.

Actual class	Output class				
	1	2	3	4	5
1	4475	95	50	222	71
2	4	305	4	9	7
3	0	17	17	3	0
4	5	1	0	77	4
5	15	0	0	36	64

Table 10: Confusion matrix of the Page-blocks dataset produced by the proposed method

Actual class	Output class				
	1	2	3	4	5
1	4913	0	0	0	0
2	0	329	0	0	0
3	0	0	28	0	0
4	0	0	0	87	0
5	0	45	0	0	70

5. Experiments

In this section, ten real-world classification experiments will be performed to illustrate the function of the method. We will provide the detailed information about the chosen dataset in the first section, and then introduce two metrics to evaluate the performance of the presented approach. Finally, some discussions are presented based on the comparison with the two related works.

5.1. Experimental datasets

In our experiments, ten datasets from the KEEL repository are used to implement the classification experiments. These datasets represent real-world problems from different industrial fields such as medical diagnosis, the manufacturing industry, the transportation industry and so on. They contain different amounts of instances varying from hundreds to thousands. The same partition is applied for each dataset: in other words, the same 5-fold cross-validation is used to divide the training and test sets. For each fold, four-fifths of the dataset is regarded as the training set, and the remaining fifth is used to test the performance of the proposed method. Table 2 summarizes the brief descriptions of those datasets, including the number of instances (v), the number of attributes (N), and the number of classes (k). We removed those instances with missing values before conducting the experiments.

5.2. Performance measures

Several standard terms have been defined for the classification problems. The classification accuracy [50] is the most widely used metric to evaluate the performance of binary classification problems. The formula for quantifying binary accuracy is:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{47}$$

where *TP = True positive*; *FP = False positive*; *TN = True negative*; *FN = False negative*.

With respect to multiclass classification problems, another metric called Cohen's kappa [48] might reflect the real classification performance of the classifier. In this study, classification accuracy and Cohen's kappa are employed as evaluation measures. Therefore, the confusion matrix needs to be determined first as in Table 3. From the confusion matrix, the Cohen's kappa is calculated as below:

$$\text{kappa} = \frac{v \sum_{z=1}^{k} \mu_{zz} - \sum_{z=1}^{k} T_{rz} T_{cz}}{v^2 - \sum_{z=1}^{k} T_{rz} T_{cz}}, \tag{48}$$

where v represents the total number of instances, k is the number of classes, and μ_{zz} is a positive number for each class, which denotes the number of test samples whose input sample category is the same as the classification result. As shown in Table 3, T_{rz} and T_{cz} are counts of rows and columns. The results for kappa range from -1 to 1, where -1 indicates total disagreement, a value of 0 indicates slight consistency and a value of 1 denotes near-perfection.

30

5.3. Discussion

The results of the comparison are summarized in Table 6 and Table 7 in terms of classification accuracy and kappa, respectively. Each bold-face data point in the two Tables is the best result for each experiment. From these data, it is interesting to draw some meaningful conclusions. In Table 6, the best results almost all come from the proposed method, except in the case of the Cleveland dataset. Even for the Cleveland dataset, the result for our method is only 0.67 % smaller than the best result, which was obtained by Qin et al. Particularly, in the case of the Vehicle dataset, our method outperforms the method of Jiang et al. by 33.56 % and is also 31.9 % higher than the result of the method of Qin et al. With respect to complex cases, such as Page-blocks and the Glass dataset, with five and seven classes, respectively, the proposed method also exhibits stable performance, with outcomes nearly 20 % higher than the method of Jiang et al. and 7 % higher than the method of Qin et al. The reason for the disparity is that the other two methods do not consider the difficult decision boundaries of multiclass cases, which leads to the weak capabilities of the two methods in cases with complex circumstances. Although the original problem is resized to a smaller scale, such as a binary classification problem (e.g., the Hepatitis and Heart datasets), our method can also more effectively handle such issues. This is just not as outstanding as the multiclass case: our method increases by nearly 3 % and 10 % versus the others.

Regarding the kappa metric, the outcomes are remarkable as well, and it also supports the competitiveness of the proposed method. From the data, we can find the most significant differences. It is a substantial improvement that the kappa of our method for Heart is almost 22 % larger than that of the method of Qin et al., though there is only a 10 % difference in the classification accuracy. The reason for this is that the classification accuracy is greatly affected by imbalanced datasets: its

value is dependent on the results of classes with more instances. Therefore, it cannot effectively characterize the classifier ability. This result shows that the predicted class label is almost exactly the same as its original class label for both classes in the imbalanced Heart dataset, which can be found in Table 9 and Table 10. From the tables, with regard to class 1, two methods can classify as many instances, but the method of Qin et al. does not perform satisfactorily for class 2. Jiang et al.'s method cannot function as well in either of the two classes. Moreover, the results are more significant for another similar situation for the Page-block dataset. It is obvious that this dataset is extremely imbalanced, where class 1 contains the most instances. Table 7 shows that the kappa of the novel method is 34 % higher than that of the method of Qin et al., while the classification accuracy is only 9 % higher. Although it does identify the most instances of class 1, it performs poorly in terms of the other four classes, which can be observed from its confusion matrix in Table 9. The novel method can not only successfully classify the instances but is less biased for imbalanced datasets, which may suppose added difficulty for the identification. These two special examples exhibit the superiority of our novel method in imbalanced dataset scenarios.

6. Conclusion

In this study, we put forward a pairwise learning-based BPA determination model called the base algorithm for D-S evidence theory. Specifically, the pairwise learning method is used to divide the original complex problem into as many pairs of easily solved binary subproblems as possible, from which the output is computed by kernel density estimation. Based on the new BPA generation algorithm, D-S combination rule, fuzzy preference relation and nondominance criterion, a new decision-making scheme is designed. This new scheme can then discriminate the pairs of subproblems

32

and achieve better decision boundaries in complex scenarios from a fuzzy preference relation. The new decision-making algorithm is designed based on the binary hypothesis pair and greatly reduces the computational complexity and boundary problems, thus enabling a wider range of application scenarios. Through ten real-world classification experiments, we have verified the efficiency and practicability of our proposed approach with respect to complex multiclass cases. We have shown that the performance of our approach is superior to those of related works in some imbalanced conditions, which indicates that any binary classification method can be used to generate the BPAs and combine them. Therefore, we predict that the use of D-S evidence theory will escalate in various industrial fields.

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Z. Cao, C.-H. Chuang, J.-K. King, C.-T. Lin, Multi-channel EEG recordings during a sustained-attention driving task, Scientific Data 6 (2019) DOI: 10.1038/s41597–019–0027–4.

[2] F. Xiao, GIQ: A generalized intelligent quality-based approach for fusing multi-source information, IEEE Transactions on Fuzzy Systems (2020) DOI: 10.1109/TFUZZ.2020.2991296.

[3] Z. Cao, W. Ding, Y.-K. Wang, F. K. Hussain, A. Al-Jumaily, C.-T. Lin, Effects of Repetitive SSVEPs on EEG Complexity using Multiscale Inherent Fuzzy Entropy, Neurocomputing (2019) DOI: 10.1016/j.neucom.2018.08.091.

[4] Z. Cao, C.-T. Lin, K.-L. Lai, L.-W. Ko, J.-T. King, K.-K. Liao, J.-L. Fuh, S.-J. Wang, Extraction of SSVEPs-based Inherent fuzzy entropy using a wearable headband EEG in migraine patients, IEEE Transactions on Fuzzy Systems (2019) DOI: 10.1109/TFUZZ.2019.2905823.

[5] H. Wang, Y. P. Fang, E. Zio, Risk assessment of an electrical power system considering the influence of traffic congestion on a hypothetical scenario of electrified transportation system in new york state, IEEE Transactions on Intelligent Transportation Systems 22 (1) (2021) 142–155. doi:10.1109/TITS.2019.2955359.

[6] Y. Deng, Uncertainty measure in evidence theory, SCIENCE CHINA Information Sciences 63 (11) (2020) 210201.

[7] F. Xiao, A distance measure for intuitionistic fuzzy sets and its application to pattern classification problems, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2019) DOI: 10.1109/TSMC.2019.2958635.

[8] M. Tang, H. Liao, E. Herrera-Viedma, C. P. Chen, W. Pedrycz, A dynamic adaptive subgroup-to-subgroup compatibility-based conflict detection and resolution model for multicriteria large-scale group decision making, IEEE Transactions on Cybernetics (2020) DOI: 10.1109/TCYB.2020.2974924.

[9] F. Xiao, On the maximum entropy negation of a complex-valued distribution, IEEE Transactions on Fuzzy Systems (2020) DOI: 10.1109/TFUZZ.2020.3016723.

[10] C. Cheng, W. Ding, F. Xiao, W. Pedrycz, A majority rule-based measure for atanassov type intuitionistic membership grades in mcdm, IEEE Transactions on Fuzzy Systems DOI: 10.1109/TFUZZ.2020.3033062.

[11] L. Pan, Y. Deng, Probability transform based on the ordered weighted averaging and entropy difference, International Journal of Computers Communications & Control 15 (4) (2020) 3743.

[12] X. Gou, H. Liao, Z. Xu, R. Min, F. Herrera, Group decision making with double hierarchy hesitant fuzzy linguistic preference relations: consistency based measures, index and repairing algorithms and decision model, Information Sciences 489 (2019) 93–112.

[13] F. Feng, W. Pedrycz, On scalar products and decomposition theorems of fuzzy soft sets., Journal of Multiple-valued Logic and Soft Computing. 25 (1) (2015) 45–80.

[14] L. Fei, Y. Feng, L. Liu, W. Mao, On intuitionistic fuzzy decision-making using soft likelihood functions, International Journal of Intelligent Systems 34 (9) (2019) 2225–2242.

[15] F. Feng, J. Cho, W. Pedrycz, H. Fujita, T. Herawan, Soft set based association rule mining, Knowledge-Based Systems 111 (2016) 268–282.

[16] F. Xiao, Evidence combination based on prospect theory for multi-sensor data fusion, ISA transactions 106 (2020) 253–261.

[17] F. Xiao, Generalized belief function in complex evidence theory, Journal of Intelligent & Fuzzy Systems 38 (4) (2020) 3665–3673.

[18] Y. Deng, Information volume of mass function, International Journal of Computers Communications & Control 15 (6) (2020) 3983. doi:https://doi.org/10.15837/ijccc.2020.6.3983.

[19] F. Xiao, Generalization of Dempster–Shafer theory: A complex mass function, Applied Intelligence 50 (2020) 3266–3275.

[20] J. Deng, Y. Deng, Information volume of fuzzy membership function, International Journal of Computers Communications & Control 16 (1) (2021) 4106. doi:https://doi.org/10.15837/ijccc.2021.1.4106.

[21] R. R. Yager, Generalized Dempster–Shafer structures, IEEE Transactions on Fuzzy Systems 27 (3) (2019) 428–435.

[22] L. Fei, Y. Feng, L. Liu, Evidence combination using OWA-based soft likelihood functions, International Journal of Intelligent Systems 34 (9) (2019) 2269–2290.

[23] R. A. Aliev, W. Pedrycz, B. Guirimov, O. H. Huseynov, Clustering method for production of z-number based if-then rules, Information Sciences 520 (2020) 155–176.

[24] W. Jiang, Y. Cao, X. Deng, A novel Z-network model based on Bayesian network and Z-number, IEEE Transactions on Fuzzy Systems (2019) DOI: 10.1109/TFUZZ.2019.2918999.

[25] Y. Tian, L. Liu, X. Mi, B. Kang, ZSLF: A new soft likelihood function based on Z-numbers and its application in expert decision system, IEEE Transactions on Fuzzy Systems (2020) DOI: 10.1109/TFUZZ.2020.2997328.

[26] F. Xiao, CEQD: A complex mass function to predict interference effects, IEEE Transactions on Cybernetics (2020) DOI: 10.1109/TCYB.2020.3040770.

[27] H. Fujita, Y.-C. Ko, A heuristic representation learning based on evidential memberships: Case study of UCI-SPECTF, International Journal of Approximate Reasoning 120. doi:10.1016/j.ijar.2020.02.002.

[28] J. Ye, J. Zhan, W. Ding, H. Fujita, A novel fuzzy rough set model with fuzzy neighborhood operators, Information Sciences 544 (2021) 266–297.

[29] C. Fu, W. Chang, S. Yang, Multiple criteria group decision making based on group satisfaction, Information Sciences 518 (2020) 309–329.

[30] W. Wang, J. Zhan, C. Zhang, Three-way decisions based multi-attribute decision making with probabilistic dominance relations, Information Sciences 559 (3) (2021) 10.1016/j.ins.2021.01.028.

[31] P. Liu, X. Zhang, W. Pedrycz, A consensus model for hesitant fuzzy linguistic group decision-making in the framework of Dempster–Shafer evidence theory, Knowledge-Based Systems 212 (2020) 106559.

[32] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, Y. He, Classifier fusion with contextual reliability evaluation, IEEE Transactions on Cybernetics 48 (5) (2018) 1605–1618.

[33] X. Xu, D. Zhang, Y. Bai, L. Chang, J. Li, Evidence reasoning rule-based classifier with uncertainty quantification, Information Sciences 516 (2020) 192–204.

[34] Z. Liu, G. Li, G. Mercier, Y. He, Q. Pan, Change detection in heterogenous remote sensing images via homogeneous pixel transformation, IEEE Transactions on Image Processing 27 (4) (2017) 1822–1834.

[35] Y. Ye, X. R. Hang, J. M. Koh, J. A. Miszczak, K. H. Cheong, N.-g. Xie, Passive network evolution promotes group welfare in complex networks, Chaos, Solitons & Fractals 130 (2020) 109464.

[36] P. Liu, M. Shen, F. Teng, B. Zhu, L. Rong, Y. Geng, Double hierarchy hesitant fuzzy linguistic entropy-based todim approach using evidential theory, Information Sciences 547 (2021) 223–243.

[37] D. Li, Y. Deng, A new correlation coefficient based on generalized information quality, IEEE Access 7 (1) (2019) 175411–175419.

[38] F. Xiao, Z. Cao, A. Jolfaei, A novel conflict measurement in decision-making and its application in fault diagnosis, IEEE Transactions on Fuzzy Systems 29 (1) (2020) 186–197.

[39] D. Suh, J. Yook, A method to determine basic probability assignment in context awareness of a moving object, International Journal of Distributed Sensor Networks 9 (8) (2013) 972641.

[40] W. Jiang, J. Zhan, D. Zhou, X. Li, A method to determine generalized basic probability assignment in the open world, Mathematical Problems in Engineering 2016.

[41] P. Xu, X. Su, S. Mahadevan, C. Li, Y. Deng, A non-parametric method to determine basic probability assignment for classification problems, Applied intelligence 41 (3) (2014) 681–693.

[42] B. Qin, F. Xiao, A non-parametric method to determine basic probability assignment based on kernel density estimation, IEEE Access 6 (2018) 73509–73519.

[43] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, The annals of mathematical statistics (1967) 325–339.

[44] G. Shafer, A mathematical theory of evidence, Vol. 42, Princeton university press, 1976.

[45] F. Liu, X. Gao, J. Zhao, Y. Deng, Generalized belief entropy and its application in identifying conflict evidence, IEEE Access 7 (1) (2019) 126625–126633.

[46] L. Pan, Y. Deng, An association coefficient of belief function and its application in target recognition system, International Journal of Intelligent Systems 35 (2020) 85–104.

[47] P. Smets, R. Kennes, The transferable belief model, Artificial intelligence 66 (2) (1994) 191–234.

[48] Z.-L. Zhang, X.-G. Luo, Y. Yu, B.-W. Yuan, J.-F. Tang, Integration of an improved dynamic ensemble selection approach to enhance one-vs-one scheme, Engineering Applications of Artificial Intelligence 74 (2018) 43 – 53.

[49] H. Garg, S. Chen, Multiattribute group decision making based on neutrality aggregation operators of q-rung orthopair fuzzy sets, Information Sciences 517 (2020) 427–447.

[50] S. U. Jan, Y. D. Lee, I. S. Koo, A distributed sensor-fault detection and diagnosis framework using machine learning, Information Sciences 547 (2021) 777–796.