



HAL
open science

Learning a credal classifier with optimized and adaptive multi-estimation for missing data imputation

Zuo-Wei Zhang, Hong-Peng Tian, Ling-Zhi Yan, Arnaud Martin, Kuang Zhou

► To cite this version:

Zuo-Wei Zhang, Hong-Peng Tian, Ling-Zhi Yan, Arnaud Martin, Kuang Zhou. Learning a credal classifier with optimized and adaptive multi-estimation for missing data imputation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, 52 (7), pp.4092-4104. hal-03271783

HAL Id: hal-03271783

<https://hal.science/hal-03271783v1>

Submitted on 27 Jun 2021

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.

Learning a credal classifier with optimized and adaptive multi-estimation for missing data imputation

Zuo-wei Zhang, Hong-peng Tian, Ling-zhi Yan, Arnaud Martin, Kuang Zhou

Abstract—The classification analysis of missing data is still a challenging task since the training patterns may be insufficient and incomplete in many fields. To train a high-performance classifier and pursue high accuracy, we learn a credal classifier based on an optimized and adaptive multi-estimation (OAME) method for missing data imputation on training and test sets. In OAME, some incomplete training patterns are estimated as multiple versions by a global optimization method thereby expanding the training set. On the other hand, the test pattern is adaptively estimated as one or multiple versions depending on the neighbors. For the test pattern with multiple versions, the corresponding outputs with different discounting factors (weights), represented by the basic belief assignments (BBAs), are fused for final credal classification based on evidence theory. The discounting factor contains two aspects: the importance and reliability factors that are used respectively to quantify the importance of the edited version itself and to represent the reliability of the classification result of the version. The effectiveness of OAME is widely validated on several real datasets and critically compared to other related methods.

Index Terms- Missing data, credal classification, evidence theory, missing values, multiple imputation.

I. INTRODUCTION

MISSING data, also known as incomplete patterns, have seriously compromised inferences from sample analysis in many fields [1]–[13]. For example, in networked control systems, the collected dataset is often not complete and the problem of missing measurements is inevitable due to dual-rate or multi-rate sampling, irregular sampling, sensor failure,

data transmission or storage error, data package dropout [12], [13]. A number of methods, based on the three types: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR), have emerged for classifying missing data [1], [7], [9], [10]. The simplest is to directly remove incomplete patterns and only classify complete ones [1], [14], while it may only be used in the cases with a few incomplete patterns.

Imputation of missing values, providing estimations by reasoning from the observed data, is often a crucial step in data analysis [3]–[5], [15]–[18]. In the mean imputation (MI) method [1], [14], for example, the missing value is replaced by the mean of all known values of that attribute. K -nearest neighbor imputation [15] (KNNI), one of the most prevalent methods, employs the neighbors to estimate the missing values. Interestingly, a locally linear approximation (LLA) approach is proposed in [16], which can provide the optimal weights of various neighbors by the locally linear reconstruction. In fuzzy c -means imputation (FCMI) [17], the query pattern is filled with cluster centers obtained by the fuzzy c -means (FCM) [19]. Besides, a fuzzy-based information decomposition (FID) method is proposed in [20], where the missing values can be estimated from the observed data with different optimized weights. Although these methods single imputation-based are still working, increasingly pieces of literature are devoted to the study of multiple estimations. Those technologies are considered to be good ways that reasonably characterize the uncertainty and imprecision caused by missing values very well and provide information complementarity for improving the accuracy of classification.

Some prevalent multiple imputation methods, also called mixed-type data imputation or multivariate imputation, of course, like Rubin’s method [14], MICE [4], MissForest [5], GAIN [21], have been developed in the past decades. In the early Rubin’s method [1], [9], [14], the missing values are imputed M times to produce M complete datasets based on an appropriate model with random variation. This has been implemented, for example, in clinical epidemiological research under the MAR assumption and provided unbiased and valid estimates of associations based on information from the available data [6]. The MICE [3], [4] can specify the conditional model for the missing data in each incomplete variable without covering the entire dataset by the multivariate model. The MissForest method [5] can handle any type of input data and make as few as possible assumptions (*e.g.* normal distributions) about structural aspects of the data. The

Manuscript received July 13, 2020; revised February 2, 2021; accepted June 14, 2021. This work was supported by the National Natural Science Foundation of China (No.61701409), the Aero Science Foundation of China (No.20182053023), and Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University (No. CX201953). This manuscript was recommended by Associate Editor XXX. (*Corresponding author: Arnaud Martin.*)

Zuo-wei Zhang is with School of Automation, Northwestern Polytechnical University, Xi’an, China, and Univ Rennes, CNRS, IRISA, DRUID, Rue E. Branly, 22300 Lannion, France. (e-mail: zhangzuowei0720@gmail.com; zuowei_zhang@mail.nwpu.edu.cn; zuozhang@univ-rennes1.fr).

Hong-peng Tian is with School of Electrical Engineering, Zhengzhou University, Luoyang, China, and the laboratory IRISA, Ecole Normale Supérieure de Rennes (ENS Rennes), France. (e-mail: tianhongpeng1220@163.com).

Ling-zhi Yan is with School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou, China. (e-mail: 598914251@qq.com).

Arnaud Martin is with Univ Rennes, CNRS, IRISA, DRUID, Rue E. Branly, 22300 Lannion, France. (e-mail: Arnaud.Martin@univ-rennes1.fr; arnaud.martin@irisa.fr).

Kuang Zhou is with School of Mathematics and Statistics, Northwestern Polytechnical University, Xi’an, China. (e-mail: kzhoumath@nwpu.edu.cn; kzhoumath@163.com).

GAIN method [21], based on the Generative Adversarial Nets (GAN) framework [22], attempts to model the distribution of the data and then makes multiple imputations by multiple draws to capture the uncertainty of the imputed values. There also exist a number of other multiple imputation methods, for instance, the literature [13] presents two multiple imputation procedures based on SAS/STAT MI and MIANALYZE. It implements popular methods for creating imputations under monotone and non-monotone (arbitrary) patterns of missing data. Interestingly, the literature [9] holds that multiple imputations are not completely superior to single estimation in any case by analyzing a large number of data with mixed-model analyses. Of course, it is consistent with our intuition. Except that, some sophisticated classifiers particularly designed for dealing with missing data without estimation have also been developed [23], [24]. This paper focuses on adaptive single or multiple imputations for each specific pattern.

The above respectable multiple imputation methods have achieved good results in some contexts, however, they have not analyzed the feasibility of multiple imputations in the training set. As one knows, the training patterns may be insufficient and incomplete in some specific fields. In such a case, it is difficult to train a high-performance classifier and pursue high classification accuracy. In addition, these methods mainly focus on characterizing the uncertainty and imprecision of missing values while ignoring how to integrate these estimations to improve the accuracy of classification. On the basis of the above analysis, we propose an optimized and adaptive multi-estimation (OAME) method for missing data imputation on training and test sets based on evidence theory.

Recently, evidence theory (ET) [25]–[28], also known as Dempster-Shafer theory or the theory of belief functions, is often employed to integrate multi-source uncertain complementary information because it can well characterize the uncertainty in the processing and has been widely used in classifier (multi-source) fusion [29]–[33]. In [31], for example, it is reported that evidence theory produced better performance than Sugeno’s fuzzy integral and the possibility theory in fusion of sources of information for target recognition. There are also some methods developed for missing data based on evidence theory [34]–[36]. For instance, an interesting adaptive imputation method for incomplete pattern classification is introduced in [34], where the incomplete pattern will be committed to a particular class if its available information is crucial for the classification. If not, K -nearest neighbor and Self-Organizing Map techniques are employed to fill multiple versions of the missing values and then an improved evidence fusion rule is used to fuse the pieces of evidence¹. In particular, the method, so-called prototype-based credal classification [36], tries to replace the missing values with the corresponding values of each class center in the training set and then to obtain different sub-versions with various estimations for the incomplete pattern. Then, the classification results of these sub-patterns with various weights are fused for decision-making. However, they do not evaluate the rationality

¹The classification result of each version is regarded as a piece of evidence for the incomplete pattern classification.

of the imputation strategies.

In this paper, we learn a credal classifier based on an optimized and adaptive multi-estimation (OAME) method for missing data imputation on training and test sets. The contributions mainly include three aspects. 1) We introduce an optimized multi-estimation strategy to impute the training set, which can effectively improve the performance of the classifier when the training patterns are seriously insufficient and incomplete. 2) We propose an adaptive imputation method to fill each incomplete pattern as single or multiple versions. This can capture the imprecision of estimations and avoid meaningless multiple estimations. 3) We design a discounting method including the importance factor and the reliability factor to evaluate the importance of the edited version and the reliability of its classification result. By doing so, we can train a high-performance classifier and pursue high accuracy. The proposed OAME mainly contains two parts: 1) Missing data imputation on training and test sets; 2) Credal classification of the query patterns.

Step 1: Some incomplete patterns in the training set are estimated as multiple versions by an optimization method while the remaining ones are filled with single-value imputation. By doing this, we can effectively expand the training set thereby improving the performance of the classifier. By contrast, each incomplete test pattern is imputed only as one single version if observed neighbors are all from the same class. Otherwise, it will be imputed as multiple versions according to classes in the neighbors. In this way, it can not only reasonably characterize the uncertainty and imprecision due to missing data but also avoid meaningless multiple estimations.

Step 2: Those test patterns that are estimated as a single version will be directly recognized by the trained classifier, while those that are estimated as multiple versions will obtain multiple corresponding classification results. These results for the pattern can be regarded as different evidence, and these pieces of evidence then are fused with different discounting (weighting) factors to make the final decision. The discounting factor contains two aspects: the importance factor and the reliability factor. The importance factor quantifies the importance of the edited version, and the reliability factor represents the reliability of the classification result of that version.

The rest of this paper is organized as follows. After a brief introduction of some related works in Section II, the new OAME method is introduced in detail in Section III. The proposed method is then tested in Section IV and compared with several other related methods, followed by conclusions.

II. RELATED WORKS

A. Basics of evidence theory

Evidence theory (ET) [25]–[28], also known as the theory of belief functions or Dempster-Shafer theory, makes it possible to model uncertainty and imprecision by defining functions on a power-set 2^Ω of a frame of discernment $\Omega = \{\omega_1, \dots, \omega_C\}$. The power-set 2^Ω is composed of all the subsets of Ω , which contains $2^{|\Omega|}$ elements. For example, if $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $2^\Omega = \{\emptyset, \omega_1, \omega_2, \omega_3, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$.

A basic belief assignment (BBA) is introduced in this theory to express the degrees of support for different elements in 2^Ω and it is a function $m(\cdot)$ from 2^Ω to $[0,1]$ satisfying:

$$\begin{cases} \sum_{A \in 2^\Omega} m(A) = 1 \\ m(\emptyset) = 0 \end{cases} \quad (1)$$

$m(A)$ denotes the extent to which evidence supports the occurrence of proposition A , but does not support the occurrence of any true subset of A . If $m(A) > 0$, A is called focal element. If $0 < m(A) < 1$, the mass of belief on A represents an uncertainty on the occurrence of A . The classification based on evidence theory allows pattern to be assigned to singleton class (e.g. ω_c , $c = 1, \dots, C$) to represent the exact information or meta-class (e.g. $\{\omega_c, \dots, \omega_{c'}\}$) composed of multiple singleton classes (i.e. $\omega_c, \dots, \omega_{c'}$, $c \neq c'$) to express imprecise information (i.e. partial ignorance). Specially, $m(\Omega)$ denotes the total ignorance degree, and it usually plays a particular neutral role in the fusion processing because it characterizes the vacuous belief source of evidence. Thus, a simple BBA given by only two focal elements ($A \in 2^\Omega$ and Ω) can represent both uncertainty and imprecision.

In classification problems, the output of the classifier can be used as an evidence to represent a BBA. Dempster's (DS) rule has been widely used in the combination of multiple sources of evidence because it is commutative and associative, which makes it more attractive in practical applications. Assuming that there are two evidences $B, C \in 2^\Omega$ where the basic belief assignments are $m_1(B)$ and $m_2(C)$, then the combination with the DS rule is defined by:

$$m(A) = \begin{cases} \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - \mathcal{K}}, & \forall A \in 2^\Omega, A \neq \emptyset \\ 0, & A = \emptyset \end{cases} \quad (2)$$

with

$$\mathcal{K} = \sum_{B \cap C = \emptyset} m_1(B)m_2(C), \quad \mathcal{K} \neq 1. \quad (3)$$

The factor \mathcal{K} can be interpreted such as a measure of the degree of conflict between the pieces of evidence. The smaller the value \mathcal{K} is, the smaller the conflict information between the pieces of evidence is. Since the rule distributes conflict information proportionally to all elements, unreasonable results may be obtained when dealing with high conflict evidence fusion. Thus, a number of alternative combination rules have emerged to overcome the limitations of DS rule, such as Yager's rule, Dubois-Prade rule, and proportional conflict redistribution (PCR6) rule [37].

In [25], Shafer proposes a discounting method for reliability evidence, which can effectively reduce the degree of conflict between evidence and discount the partial mass of belief in a BBA to the total ignorance Ω according to the reliability (weight) factor, defined as:

$$\begin{cases} \alpha m(A) = \alpha \cdot m(A), A \subset \Omega, A \neq \Omega \\ \alpha m(\Omega) = 1 - \alpha + \alpha \cdot m(\Omega) \end{cases} \quad (4)$$

where α is the reliability factor and $\alpha \in [0,1]$. The OAME

employs this method as the basis for discounting different evidence. The DS rule is then used to combine the classification results provided by different versions because its associativity property makes it easier to implement than other fusion rules.

Particularly, the well-known pignistic probability transformation method [26], which approximates a BBA to probability measure, is usually used for decision-making in ET, and the pignistic probability of the singleton class ω_c is defined by:

$$BetP(\omega_c) = \sum_{A \in 2^\Omega, \omega_c \in A} \frac{1}{|A|} m(A) \quad (5)$$

where $|A|$ denotes the cardinality of A , i.e. the number of singleton classes included in A .

B. Related methods

Since some related representative methods have been developed and employed as the comparisons, we briefly review these methods in this subsection and clarify the differences between the proposed method and these existing works. Specifically, the respectable methods are KNNI [15], FCMI [17], MissForest (MF) [5], GAIN [21], PCC [36], CCAI [34], FID [20], LLA [16] and IVIACLR [18].

First, the classical KNNI [15] and FCMI [17] methods are employed as the representatives of classical single estimation strategies. The KNNI [15] estimates missing values using the average of the corresponding attributes of the neighbors. In FCMI [17], the dataset is first clustered by FCM, then the cluster centers and the membership degrees of the incomplete pattern are employed to estimate its missing values. Besides, the single imputation methods presented in recent years, named FID [20], LLA [16] and IVIACLR [18], are also employed as competitions. The FID method fills the incomplete pattern from the observed data with different optimized weights obtained by the fuzzy membership functions, which takes into account different contribution of the observed data. The neighbors with optimal weights obtained by the locally linear reconstruction are used to yield optimal estimation for the incomplete patterns in LLA. In IVIACLR, a simple imputation method (e.g. MI) is employed to initial imputations, then the missing values are filled by clusterwise linear regression technology.

Then, we employ MissForest (MF) [5] and GAIN [21] as the benchmarks for multiple imputations. The MF method provides multiple estimations for the incomplete pattern by averaging over many unpruned classification or regression trees. In GAIN, the goal of the generator (G) is to accurately fill the incomplete pattern and the discriminator (D) aims to distinguish whether the value is the estimation or actually observed, and some additional information in the form of a hint vector provides for D to obtain multiple estimations and ensure G models the distribution of the data.

Finally, the methods called PCC [36] and CCAI [34] integrate multiple imputations and credal classification based on evidence theory. Here we explain the difference between OAME and them. In PCC, it assumes that the training set is complete and sufficient, which may be out of line with reality sometimes. OAME can fill training patterns and effectively

improve the performance of the classifier by an optimization method, especially when the training patterns are seriously insufficient and incomplete. In addition, PCC fills the incomplete test pattern by each class center in the training set, which is not accurate enough and makes some meaningless multiple estimations. In OAME, the test patterns are adaptively imputed as single or multiple versions based on the observed neighbors to overcome the shortcomings of PCC. Lastly, PCC just considers the reliability of classification results of evidence while OAME not only evaluates the reliability of evidence but also the importance of the edited version. In CCAI, an adaptive imputation method is proposed for incomplete pattern classification. In the first step of CCAI, the training set is assumed complete and the incomplete pattern in the test set will be committed to a particular class if its available information is crucial for the classification, which mainly takes into account the necessity of estimation, whereas the OAME method is devoted to expanding the patterns using an optimal estimation strategy in the training set. In the second step, CCAI uses the K nearest weighting vectors of each class respectively to estimate multiple estimations for missing values. Then the discounted classification results of these sub-patterns with various estimations are fused by an improved evidence fusion rule for the final decision. However, OAME mainly focuses on the diversity of importance and reliability for the pieces of evidence. A summary of these methods is given in Section E of the supplementary file to describe their features, advantages and disadvantages more intuitively.

III. OPTIMIZED AND ADAPTIVE MULTI-ESTIMATION FOR CREDAL CLASSIFICATION

In this section, the proposed OAME method is introduced in detail to address the problem that trains a high-performance classifier by extending the training set and pursues high classification accuracy by integrating pieces of estimations. It consists of two parts: 1) Missing data imputation on training and test sets; 2) Credal classification of the query patterns. These two parts will be discussed in subsections III-A and III-B, respectively. In subsection III-C, the tuning of involved parameters will be introduced. The analysis of complexity is included in III-D. In addition, we list all the notations in this paper and give a brief introduction of them in Section E of the supplementary file for convenience.

A. Missing data imputation on training and test sets

Since the training patterns with missing values have the label information, we must take a very cautious multi-estimation strategy to expand the training set. As a simple example, Fig. 1 shows a 3-class problem with two dimensions of attributes corresponding to x -coordinate and y -coordinate. If the training pattern O_2 loses the x -dimensional attribute, for example, one can reasonably obtain the neighbors (*i.e.* O_1 , O_3 , O_5 and O_6) from different classes (*i.e.* $O_1, O_3 \in \omega_1$ and $O_5, O_6 \in \omega_2$) depending only on the y -dimensional attribute. In such a case, the incomplete pattern O_2 has the potential to be estimated as two versions, *i.e.* O'_2 with estimation belonging to ω_1 and O''_2 with estimation belonging to ω_2 . O'_2 can be considered

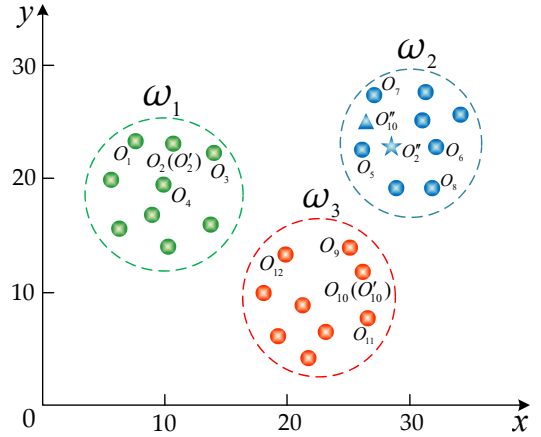


Figure 1: Illustration of feasibility analysis of multi-estimation for the training patterns.

as the complete pattern O_2 itself and O'_2 represented by the pentagram is an expanded training pattern in ω_2 . By doing this, we can effectively expand the training set. If the training pattern O_2 loses the y -dimensional attribute, however, one can only impute once for the missing value x since the neighbors are all from ω_1 . One can also get similar inference from the pattern O_{10} when y losses if interesting, where O''_{10} indicated by the triangle is an expanded training pattern.

From the above analysis, one can find that those patterns that meet the following two conditions simultaneously have the right to be estimated as multiple versions in applications: 1) The pattern loses multiple attribute values; 2) The neighbors come from different classes. The condition 1) is to prevent the method of measuring similarity from being inaccurate, which can be regarded as a prudent strategy. Assuming that the training pattern loses only one attribute value, we think that it is not prudent to take multi-estimation for the pattern even though the neighbors come from different classes since we cannot judge whether the neighbors are true, especially for high-dimensional data. Therefore, the key is which incomplete training patterns should be estimated as multiple versions? Here we assume that condition 1) is satisfied when the number of missing values is not less than the threshold δ . The multi-estimation processing is then explained in detail as follows.

Let us consider a test set $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_H\}$ and classify it with a training set $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ including s attributes in the class editing framework $\Omega = \{\omega_1, \dots, \omega_C\}$. Assuming that \mathcal{X} includes two parts: the set \mathcal{X}_1 with complete patterns and the set \mathcal{X}_2 with incomplete patterns, *i.e.* $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2$. For a specific pattern, say $\mathbf{x} \in \mathcal{X}_2$ with δ' ($\delta' < s$) missing values, we consider that its known attributes do not play a decisive role in determining the class if δ' exceeds the threshold δ . In other words, the class label is little significance with $\delta \leq \delta'$ since the larger the δ' is, the more attributes lose². In such a case, the pattern \mathbf{x} has the potential to be estimated as multiple versions with different classes to extend the training set.

Here we assume that the known attributes of \mathbf{x} are true and reliable, and the neighbors, $\mathbf{x}_1, \dots, \mathbf{x}_K$, belonging to

²Here we consider that the more attributes the pattern lose, the more indistinct its class is.

\mathcal{X}_1 contain G ($1 \leq G \leq \mathcal{C}$) class information³, denoted as $\omega_g, g = 1, \dots, G$. In such a case, the pattern \mathbf{x} should be edited as G versions, say \mathbf{x}^g , to characterize the uncertainty caused by missing values and also to extend the training set. Mathematically, the g -th version of \mathbf{x} is computed for $g = 1, \dots, G$ by:

$$\mathbf{x}^g = \sum_{k=1}^K \lambda_k \cdot \mathbf{x}_k^g \quad (6)$$

where \mathbf{x}_k^g is the k -th neighbor of \mathbf{x} in class ω_g , and λ_k denotes the weight of \mathbf{x}_k^g . K is the number of neighbors. Since the distance between the neighbors and \mathbf{x} are different, the importance of these neighbors is also various. A commonly used and effective method is employed here to calculate the weight λ_k and denoted as follows:

$$\lambda_k = \frac{\exp(-\|\mathbf{x} - \mathbf{x}_k^g\|)}{\sum_{k=1}^K \exp(-\|\mathbf{x} - \mathbf{x}_k^g\|)} \quad (7)$$

where $\|\cdot\|$ represents the Euclidean distance⁴. From Eq. (7), one can see that the nearer the neighbor to \mathbf{x} , the larger the weight obtained, which means that the more important it is and the greater the contribution it will make in the process of estimating the missing values. By doing this, one can also get the other $G - 1$ versions of \mathbf{x} in the corresponding classes using Eqs. (6) and (7), *i.e.* the pattern \mathbf{x} will be eventually edited as G versions in total to extend the training set.

The missing values of \mathbf{x} , however, are estimated into multiple versions that may bring a greater risk of error imputation if δ' is less than the threshold δ . In such a case, a more reliable single imputation strategy should be used. Specifically, the missing values will be imputed once with the neighbors from the same class. Based on the above principles, one can easily find that too small δ will lead to only a few patterns that are used for single imputation while a large number of patterns that are used for multiple imputations. This may increase the risk of error imputation. By contrast, if most incomplete patterns adopt single imputation in the class, it can reduce the risk of error, however, this cannot effectively expand the training set. Thus, we need to make a compromise between single and multiple imputations by the optimized $\hat{\delta}$. Next, we will explain in detail how to optimize δ .

We can obtain multiple corresponding extended training sets under the different threshold δ . The optimization process for threshold δ is based on a basic assumption. The extended training set that makes the performance of the basic classifier reach the optimal level is the best one and the corresponding threshold $\hat{\delta}$ at this time is the optimized value. The performance index of the basic classifier is defined as the difference between the classification results of the patterns in the complete training set \mathcal{X}_1 and the real labels of the patterns. The smaller the difference, the better the performance.

Assuming that one can get a corresponding training set \mathcal{X}^δ

³Although we consider that $G = 1$ is possible, in fact, the possibility is very small since the larger δ' , the more classes the neighbors may contain.

⁴In very high dimensional space, it may be unreliable to use Euclidean distance to measure the similarity. In such a case, other distances (*e.g.* Fractal distance) can be used.

for any δ ($\delta \in [1, s]$). Then, one can use the basic classifier trained by \mathcal{X}^δ to classify the \mathcal{X}_1 with N' patterns, $N' < N$. By convention, the real class (known) of the pattern, say \mathbf{x}_n , is denoted by $L(n)$, and its estimated class declared by the classifier is denoted $\hat{L}(n)$. For instance, $L(n) = [0, 1, 0]^T$ if \mathbf{x}_n belongs to the class ω_2 under the frame $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The classification result of \mathbf{x}_n , obtained by the classifier $\Gamma(\cdot)$, is denoted by $\boldsymbol{\mu}_n = [\mu_n(1), \dots, \mu_n(\mathcal{C})]^T$. Each component $\mu_n(c)$ represents the probability of the pattern associated with class ω_c , *i.e.* $\mu_n(c) \triangleq P(\hat{L}(n) = \omega_c)$, $c = 1, \dots, \mathcal{C}$. If the pattern truly belongs to class $\omega_{c'}$, the perfect classifier output for the pattern should be consistent with the truth as $\boldsymbol{\mu}_n : \mu_n(c') = 1$ and $\mu_n(c) = 0$ for $c \neq c'$. However, the classifier output is usually more or less different from the truth in practice, and the classification result of the n -th pattern in \mathcal{X}_1 can be defined as follow:

$$\boldsymbol{\mu}_n = \Gamma(\mathbf{x}_n | \mathcal{X}^\delta) \quad (8)$$

where $\Gamma(\cdot)$ represents the standard (basic) classifier and the choice of $\Gamma(\cdot)$ is left to one's preference. For instance, K-NN [38], EK-NN [39], or SVM [40] can be employed here. It is expected that the results obtained by $\Gamma(\cdot)$ with \mathcal{X}^δ are as close to the ground truth as possible in \mathcal{X}_1 , which is expressed as the set of following equations:

$$\begin{cases} \|L(1) - \boldsymbol{\mu}_1\| = \epsilon_1 \\ \|L(2) - \boldsymbol{\mu}_2\| = \epsilon_2 \\ \vdots \\ \|L(N') - \boldsymbol{\mu}_{N'}\| = \epsilon_{N'} \end{cases} \quad (9)$$

The global criterion \mathcal{O}_δ for the threshold δ can be denoted as the sum of all errors as follows:

$$\mathcal{O}_\delta = \frac{1}{N'} \sum_{n=1}^{N'} \epsilon_n = \frac{1}{N'} \sum_{n=1}^{N'} \|L(n) - \boldsymbol{\mu}_n\| \quad (10)$$

One can find that different δ corresponds to different \mathcal{X}^δ and produces various \mathcal{O}_δ . For a high-dimensional dataset, we can randomly select several δ_i , and then fit a curve function $f(\cdot)$ defined as follows:

$$f(\delta) = \mathcal{F}(\delta_i, \mathcal{O}_{\delta_i}) \quad (11)$$

where $\mathcal{F}(\cdot)$ represents the fitting method, and the least square curve fitting [41] is employed here in this paper. Therefore, the optimal $\hat{\delta}$ is defined as the integer value of δ satisfying the minimum $f(\delta)$, and denoted mathematically for $\delta_i = 1, \dots, s - 1$ as follows:

$$\hat{\delta} = [\delta]$$

with

$$f(\delta) \leq f(\delta_i) \quad (12)$$

where $[\cdot]$ is the rounding symbol. Of course, for general s , one can also let δ_i directly take all possible values and get the optimal $\hat{\delta}$ by comparing the values of different \mathcal{O}_{δ_i} . By doing this, one can obtain the optimized training set $\mathcal{X}^{\hat{\delta}}$ to train the basic classifier $\Gamma(\cdot)$.

In the test set \mathcal{Y} , we propose an adaptive multiple imputation method for incomplete patterns, which is similar to imputing

x. The main procedure is to use the complete patterns in the training set to estimate the missing values of the query pattern \mathbf{y} . For the query (incomplete) pattern \mathbf{y} , let us consider that there are G' ($1 \leq G' \leq \mathcal{C}$) classes in the neighbors. Here we choose the patterns in \mathcal{X}_1 as the neighbors to avoid the possible negative effects of the estimations in the training set \mathcal{X}^δ . Thus, the query pattern \mathbf{y} can be edited as G' versions using Eqs. (6) and (7), named $\mathbf{y}^1, \dots, \mathbf{y}^{G'}$, to model the uncertainty caused by the missing values.

B. Credal classification of the query patterns

The edited test set \mathcal{Y} thereby is classified by the chosen classifier $\Gamma(\cdot)$ depending on the dataset \mathcal{X}^δ . Here the outputs can be a Bayesian BBA if the chosen classifier works under probability framework (e.g. K-NN [38] or SVM [40]), and it can also be a regular BBA having some mass of beliefs committed to the ignorant class Ω if the classifier works under evidence theory (e.g. EK-NN [39]). For the test pattern \mathbf{y} estimated as a single version, it can be directly recognized by the classifier $\Gamma(\cdot)$. For the test pattern \mathbf{y} with multiple versions, by contrast, one can obtain G' classification results, and those results are used together to classify the pattern \mathbf{y} .

Evidence theory has been reported that able to deal well with uncertain and imprecise information and produce better performance in the fusion of sources of information for pattern recognition [31]. In such a case, these outputs of corresponding versions, $\mathbf{y}^1, \dots, \mathbf{y}^{G'}$, regarded as complementary information, are integrated on the basis of evidence theory to make the final decision for \mathbf{y} . Here each output can be regarded as one evidence represented by the basic belief assignments (BBAs). One should note that, however, the information used to estimate the multiple versions are inconsistent. Thus, the importance of different versions are naturally various for \mathbf{y} . In addition, the classifier obtained by global optimization with \mathcal{X}^δ may not be suitable for each special query pattern⁵. Thus, the reliability of classification results is also various.

Here the parameters $\alpha^{g'}$ and $\beta^{g'}$, $g' = 1, \dots, G'$, are employed to represent the importance factor and the reliability factor, respectively. The importance factor $\alpha^{g'}$ is used to quantify the importance of the edited version itself, and the reliability factor $\beta^{g'}$ is used to represent the reliability of the output of the version. Next we introduce how to obtain them.

1) The importance factor $\alpha^{g'}$. The neighbors from \mathcal{X}_2 generally contain multiple classes. The distances between these patterns and \mathbf{y} are different, and multiple patterns may belong to the same class. In general, if multiple patterns come from the same class $\omega_{g'}$ and these patterns are closer to \mathbf{y} , the version $\mathbf{y}^{g'} \in \omega_{g'}$ is more important for the query pattern \mathbf{y} . A common distance-based method is employed here to get

the importance factor $\alpha^{g'}$ for the version $\mathbf{y}^{g'}$, and defined by:

$$\alpha^{g'} = \frac{\sum_{\mathbf{x}_k \in \omega_{g'}} \exp(-\|\mathbf{y} - \mathbf{x}_k\|)}{\sum_{k=1}^K \exp(-\|\mathbf{y} - \mathbf{x}_k\|)} \quad (13)$$

By doing this, one can obtain the importance factor $\alpha^{g'}$. The importance factor, however, cannot replace the reliability of the output for $\mathbf{y}^{g'}$ since the performance of the classifier is often constrained by many factors [42]. For example, the training patterns of some classes are not sufficient or the globally optimized classifier is not applicable to each specific pattern.

2) The reliability factor $\beta^{g'}$. Since the neighbors play an important role in estimating missing values and classification, we employ the neighbors to provide (more or less) supplementary information here. As one knows, if the query pattern \mathbf{y} originally belongs to $\omega_{g'}$, then $\mathbf{y}^{g'}$ should be the most reliable version and the classification result of $\mathbf{y}^{g'}$ should also be the most reliable for \mathbf{y} . In such a case, the pattern $\mathbf{y}^{g'}$ and the neighbors in $\omega_{g'}$ are the most similar because it is estimated from these neighbors⁶. This also means that the classification results of $\mathbf{y}^{g'}$ and the neighbors are the most similar regardless of the performance of the basic classifier $\Gamma(\cdot)$. In other words, the more similar the classification results between $\mathbf{y}^{g'}$ and the neighbors in the same class $\omega_{g'}$, the more reliable the result of $\mathbf{y}^{g'}$ is. Thus, we employ the neighbors to help us quantify the reliability of different classification results for \mathbf{y} . Here we consider the distance $d_{g'}$ of the classification results between the pattern $\mathbf{y}^{g'}$ and the neighbors (e.g. $\mathbf{x}_k^{g'}$) in $\omega_{g'}$ as the criterion to measure similarity, and defined by:

$$d_{g'} = \sum_{k=1}^K \lambda_k \cdot \|\boldsymbol{\mu}_{g'} - \boldsymbol{\mu}_k\| \quad (14)$$

where $\boldsymbol{\mu}_{g'}$ (or $\boldsymbol{\mu}_k$) is the classification result of $\mathbf{y}^{g'}$ (or the neighbor $\mathbf{x}_k^{g'}$) obtained by $\Gamma(\cdot)$, and λ_k is the weight obtained by Eq. (7) and has been explained before. Thus, the reliability factor $\beta^{g'}$ is defined by:

$$\beta^{g'} = \frac{\exp(-d_{g'})}{\sum_{g'=1}^{G'} \exp(-d_{g'})} \quad (15)$$

where $\beta^{g'}$ is the reliability factor of the output for $\mathbf{y}^{g'}$. To more intuitively show the basic principles and acquisition process of the importance and reliability factors $\alpha^{g'}$ and $\beta^{g'}$, we give specific examples in Section A of the supplementary file. After obtaining the importance factor $\alpha^{g'}$ and the reliability factor $\beta^{g'}$, the Shafer's discounting method is applied here to discount the pieces of evidence for \mathbf{y} , called the discounting factor $\gamma^{g'}$ for $g' = 1, \dots, G'$, since the G' classification results

⁵In our recent work, a new pattern classification accuracy improvement (CIA) method working with local quality matrix is proposed to overcome the shortcomings of global optimal classifier. If interesting, one can refer to the literature [42].

⁶It should be noted that the reliable classification result does not mean that the performance of the classifier is also reliable at this time, because the performance is often constrained by many factors and also related to the distribution of $\mathbf{y}^{g'}$. That is the reason why we employ the classification results of the neighbors to quantify the reliability of the classification result for $\mathbf{y}^{g'}$.

may strongly support different classes⁷. The discounted masses of belief are obtained by:

$$\begin{cases} m_{g'}(A) = \gamma^{g'} \mu_{g'}(A), A \subset \Omega; \\ m_{g'}(\Omega) = 1 - \gamma^{g'} + \gamma^{g'} \mu_{g'}(\Omega). \end{cases} \quad (16)$$

subject to

$$\gamma^{g'} = \frac{\alpha^{g'} \beta^{g'}}{\sum_{g'=1}^{G'} \alpha^{g'} \beta^{g'}} \quad (17)$$

where $\mu_{g'}(A)$ represents the possibility that pattern $\mathbf{y}^{g'}$ belongs to the focal element A , and $\mu_{g'}(A) = \mu_{g'}(c)$ if $A = \omega_c$. By doing this, one can convert the probability outputs, if the basic classifier works with the probability framework, to the mass of beliefs thanks to the Shafer's discounting method [25]. Whereas if the basic classifier (*e.g.* EK-NN [39]) works with belief functions, the focal element A may also represent the total unknown class Ω . Therefore, $\mu_{g'}$ can represent probability or mass of beliefs, which is determined by the chosen basic classifier. One can also find that the more important and reliable the classification result is, the larger the corresponding discounting factor is, and the less discounted information assigned to the total unknown class Ω is.

The degree of conflict between discounted BBAs is less intense because the discounted information is assigned to unknown class (*i.e.* Ω). In such a case, the famous DS rule, defined in Eqs. (2) and (3), can be used to integrate these discounted BBAs. Then, the fused BBAs of the pattern \mathbf{y} are transferred into pignistic probability by Eq. (5) for the final decision of \mathbf{y} .

The pseudo-code of OAME is presented in **Algorithm 1** to clearly show how OAME works.

C. Tuning of parameters

In the proposed (OAME) method, the numbers of neighbors should be tuned in the real applications, which contains two parts: the number of global neighbors, called \mathcal{K}_1 , and the number of local neighbors, called \mathcal{K}_2 .

The \mathcal{K}_1 neighbors are used to assist the query (training or test) pattern (*i.e.* \mathbf{x} or \mathbf{y}) in determining which classes it may belong to, and also used to estimate the importance factor $\alpha^{g'}$ of different versions $\mathbf{y}^{g'}$ for the query pattern \mathbf{y} . Since they come from the complete dataset \mathcal{X}_2 including patterns from all classes, the \mathcal{K}_1 should not be too small because it may cause the neighbors found for the query pattern to not theoretically cover all classes. If \mathcal{K}_1 is too large, however, it is possible to involve unrelated classes in the discussion, which is not helpful to estimate the possible versions of the query pattern (\mathbf{x} or \mathbf{y}). Thus, we recommend that \mathcal{K}_1 should not be valued less than the class number \mathcal{C} in the dataset (*i.e.* $\mathcal{K}_1 \geq \mathcal{C}$) in general and $\mathcal{K}_1 \in [8, 15]$ can be used as the default in most cases depending on our experiences.

⁷In this case, if the DS rule is used to fuse them directly, one often gets an unreasonable result. This is also the reason why we employ the Shafer's discounting method here.

Algorithm 1 Optimized and adaptive multi-estimation method for missing data credal classification.

Require: The training set: $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$; The test set: $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_H\}$; The chosen classifier: $\Gamma(\cdot)$; The given parameters: $\mathcal{K}_1, \mathcal{K}_2$.

Ensure: Class decision results.

for $\delta = 1$ to $s - 1$

for $n = 1$ to N

 Select the neighbors for \mathbf{x} ;

 Impute \mathbf{x} into G' version using Eqs. (6)-(7);

end

end

Classify \mathcal{X}^δ with estimations using Eq. (8);

Optimize the $\hat{\delta}$ using Eqs. (9)-(12).

for $h = 1$ to H

 Impute \mathbf{y} into G' version using Eqs. (6)-(7);

If $G' = 1$

 Output the classification of \mathbf{y} for decision-making.

else

 Calculate the importance factor $\alpha^{g'}$ using Eq. (13);

 Calculate the reliability factor $\beta^{g'}$ using Eqs. (14)-(15);

 Discount the classification results using Eqs. (16)-(17);

 Fuse the discounted results using Eqs. (2)-(3);

 Make decision for \mathbf{y} using Eq. (5).

end

end

return Class label.

The \mathcal{K}_2 neighbors are used to estimate the missing values of the query pattern ($\mathbf{x}^{g'}$ or $\mathbf{y}^{g'}$) in the specific class (ω_g or $\omega_{g'}$), and also used to assist the specific version $\mathbf{y}^{g'}$ in estimating the reliability factor $\beta^{g'}$. Since the neighbors come from the specific class (ω_g or $\omega_{g'}$), \mathcal{K}_2 does not need to be too large and it is an open value. Thus, we recommend a common value, *i.e.* $\mathcal{K}_2 = 5$, as the default.

D. Complexity analysis

The complexity of OAME is analyzed here. Let us consider that there are H incomplete patterns in the test set and they are classified using the training set consisted of N_1 complete patterns and N_2 incomplete patterns under the class editing framework $\Omega = \{\omega_1, \omega_2, \dots, \omega_{\mathcal{C}}\}$. The computational complexity of OAME mainly comes from the calculation of Euclidean distances between patterns to find neighbors for filling incomplete patterns. In the training set, each incomplete pattern needs to calculate N_1 distances to search \mathcal{K}_1 neighbors that contain G_i class information. Then it needs to find \mathcal{K}_2 neighbors in these G_i classes to estimate missing values, respectively. Thus, the computational complexity of the incomplete patterns in training set is $\mathcal{O}(N_1 N_2) (\sum_{i=1}^{N_2} \sum_{c=1}^{G_i} N_1^c)$, where N_1^c represents the number of complete training patterns in class ω_c . In the test set, the principle of estimating missing values for the H incomplete patterns is similar to that of filling in the incomplete training patterns, so that the computational complexity of the test pattern can be obtained

as $\mathcal{O}(N_1 H) \left(\sum_{i=1}^H \sum_{c=1}^{G_i} N_1^c \right)$. Therefore, the total computational complexity of OAME is $\mathcal{O}(N_1(N_2 + H) \left(\sum_{i=1}^{N_2+H} \sum_{c=1}^{G_i} N_1^c \right))$.

E. Some discussions on applications

Some discussions are opened in this part for the applications of the proposed OAME method and they are given in Section B of the supplementary file.

IV. EXPERIMENT APPLICATIONS

In this section, we perform extensive experiments to evaluate the performance of the proposed OAME method and compare it with several other related methods. The code of MF and GAIN are implemented in R and Python, respectively, and the code of the other methods are implemented in MATLAB. All experiments are performed on a PC with Intel Core i7-9750H CPU and 16GB RAM. The model performance is measured here by the four commonly used indexes [43]: 1) accuracy (AC); 2) precision (PE); 3) recall (RE); 4) f1-measure (F_1). Let us consider that the different classes of classification results of a test set \mathcal{Y} with H patterns is defined as $T = \{T_1, T_2, \dots, T_C\}$ and $L = \{L_1, L_2, \dots, L_C\}$ is the corresponding ground truth of the set \mathcal{Y} , where T_c ($c = 1, \dots, C$) represents the set of patterns that are assigned into class ω_c and the set of original patterns included in class ω_c is defined as L_c . The number n_c ($c = 1, \dots, C$) with $n_c = |T_c \cap L_c|$ is the number of correctly classified patterns in class ω_c . Then, these indexes are given by:

$$(1) AC = \frac{1}{H} \sum_{c=1}^C n_c; \quad (2) PE = \frac{1}{C} \sum_{c=1}^C \frac{n_c}{|T_c|};$$

$$(3) RE = \frac{1}{C} \sum_{c=1}^C \frac{n_c}{|L_c|}; \quad (4) F_1 = \frac{1}{C} \sum_{c=1}^C \frac{2n_c}{|L_c| + |T_c|}.$$

The higher the values of AC , PE , RE and F_1 , the better the performance of the classification method. In addition, the confidence interval [44] based on statistical properties, which has been demonstrated to be powerful to compare different methods, are used as the benchmark for these indicators. In experiments, we assume that for each test pattern, the probability that it is correctly classified follows Gaussian distribution. In other words, the classification result of each test pattern can be considered as an independent result. In such a case, the test set can be consider as running approximately $10 \times H$ times to calculate the confidence interval, where H is the number of test patterns and 10 is the number of executions of the test set. The confidence intervals with the confidence level as 95% for different indicators of the real datasets by different methods are shown in the corresponding Tables.

A. Benchmark datasets

Ten commonly used real datasets from the UCI repository (available at <http://archive.ics.uci.edu/ml>) are employed here to test and evaluate the performance of different methods. The basic information of these datasets including the number of classes (#Class.), attributes (#Attr.) and instances (#Inst.) are

reported in Table I. The Vehicle dataset including 846 patterns is employed here since 100 patterns are retained and not provided, and the Bank dataset originally contains incomplete patterns in the UCI repository.

Table I: Basic information of the UCI datasets.

Data	#Class	#Attr.	#Inst.
Bank (Ba)	2	19	539
Contraceptive (Con)	3	9	1473
Biodeg (Bi)	2	41	1055
Vehicle (Ve)	4	18	846
Segment (Seg)	7	19	2310
Sensor (Se)	4	24	5456
Connectionist (Co)	11	10	990
Wifi (Wi)	2	7	2000
Yeast (Ye)	3	8	1136
Parkinson (Pa)	21	21	2928

B. Performance evaluation

In order to verify the effectiveness of the proposed OAME method from different perspectives, we conduct a large number of experiments: 1) Fig. 4 is used to reveal how to obtain the optimized threshold $\hat{\delta}$, and 2) Table III and the results shown in Section C of the supplementary file are employed to report the effect of the chosen basic classifiers K-NN, EK-NN and SVM; 3) Fig. 5 and Table IV show the effect of the number (*i.e.* ℓ) of attribute losses on the performance of different methods (take the Connectionist dataset as an example), and 4) Table V is used to verify that the OAME is still more robust than other methods at different missing rates (*i.e.* 50%, 60%, and 70%).

In the process, parameters of different methods are presented in Table II, where FID has no parameters. Since the accuracy will be used as an indicator to evaluate the performance of these methods, we need to adjust the threshold (*i.e.* $\epsilon = 0$) so that PCC can no longer obtain imprecision rate in classification results.

Table II: Parameters of different methods.

Method	Parameters
KNNI	$K = 5$
FCMI	$\epsilon = 10^{-5}, m = 2$
MF	$Maxiter = 10, Tree = 100$
PCC	$\epsilon = 0$
CCAI	$\epsilon = 0, \eta = 0.7, M = 3, N = 4$
LLA	$Maxiter = 100, Tolerance = 10^{-4}$
GAIN	$\alpha = 0.1$
IVIACLRL	$O_{max} = 5, l = 5, l_{max} = 150$
OAME	$\mathcal{K}_1 = 11, \mathcal{K}_2 = 5$

The dataset Bank originally contains incomplete patterns that are used as test sets, and its complete patterns are used as training sets. The remaining nine datasets have no missing values, where half of the patterns are randomly selected as test sets and the others as training sets. In order to simulate the actual situation, 65% of the patterns in the training set will be randomly lost with $m \in [1, s)$ missing values, where s is the dimension of the complete patterns, and each pattern of the test set is randomly lost ℓ attributes [36].

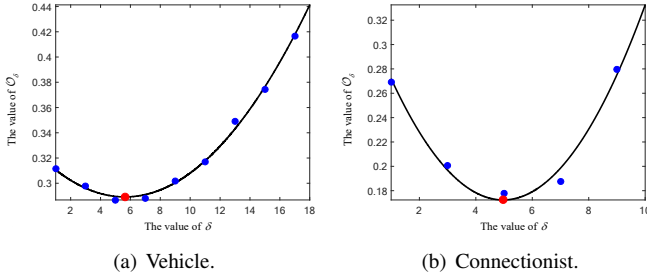


Figure 2: Fitting the curve function $f(\cdot)$ and adaptive the optimized threshold $\hat{\delta}$.

In order to visualize the optimized threshold $\hat{\delta}$, the OAME selects some values for δ (*i.e.* $\delta = 1, 3, \dots, s-1$) to obtain the corresponding average deviation \mathcal{O}_δ based on some of the employed sets, as shown in Fig. 2, where the number of missing values in incomplete training patterns corresponds to the x -coordinate and the y -coordinate represents the corresponding average deviation \mathcal{O}_δ . The fitting curves of other datasets are given in Section C of the supplementary file. Here the curve function $f(\cdot)$ depending on the least square rule [41] is used to get the minimum \mathcal{O}_δ , which corresponds to the optimized threshold $\hat{\delta}$ (*i.e.* the red dot). One can see from Fig. 2 that the minimum points of \mathcal{O}_δ are not at the two endpoints of the curve, and most of them are distributed in the relative central position, which proves that it is feasible to find such an optimized threshold $\hat{\delta}$ in OAME.

For fairness and reproducibility, each employed dataset is processed ten times randomly, then the ten different groups of each dataset are tested ten times by different methods, respectively, so as to avoid the negative effect of the accidental. The value of indicators is the average of all the results for the dataset, which makes our experimental results more reliable.

The confidence intervals of accuracy for different methods with K-NN are given in Table III. The corresponding results of EK-NN and SVM are shown in Section C of the supplementary file. The average of confidence intervals (Ave) on different datasets with the same classifier is given in the second last row to show the general performance of the various method and the winning times (WT)⁸ of different methods on these datasets is shown in the last row. From the results, one can see that the confidence intervals of accuracy of OAME method is significantly higher than that of comparison methods since the OAME method effectively improves the performance of the classifier by expanding the number of patterns in the training sets. In addition, the results of EK-NN and SVM show that the performance of OAME is still better than that of other methods even different classifiers are applied, which indicates that the implementation of OAME does not depend on the selection of the basic classifiers.

Since we only show partial results in Table III when ℓ takes different values, thus, a dataset, named Connectionist with different ℓ , is employed here to avoid the negative impact of the randomness of ℓ -value selection and study the impact of different ℓ values on classification performances of different

⁸One method will win one time if it produces the maximum accuracy rate for one classification case with respect to the other methods.

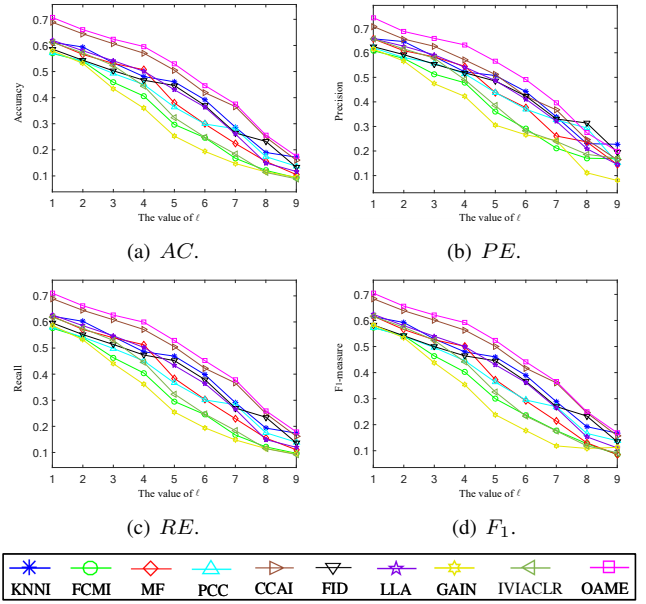


Figure 3: The classification results for Connectionist dataset with various ℓ by different methods.

methods. The values of the indexes (AC, PE, RE, F_1) of various methods in classifying the Connectionist dataset with different ℓ using K-NN as shown in Fig. 3, where the x -coordinate denotes the different ℓ values, ranging from 1 to 9, and the y -coordinate represents the values of different indexes, which are expressed in $[0,1]$. In addition, the average classification results for Connectionist dataset with various ℓ by different methods are given in Table IV. The results intuitively reveal that the performance of OAME is still better than that of other methods even if the different number of missing values is set. Additionally, with the increase of the number (*i.e.* ℓ) of missing values in the test set, the performance of different methods tends to be worse since the more missing data, the greater the deviation of the estimations, which may increase the risk of misclassification of incomplete patterns.

In the previous experiment, we only tested the case where the missing rate was 65% in the training set, and also test the performance of different methods in the case of the different numbers of missing values in the test set. Here, the incomplete patterns of the dataset Bank that original contains missing values makes up the test set and the complete patterns are used as training set, of which 50%, 60%, and 70% training patterns randomly lose the missing values. The classification results of different methods are given in Table V and the results of other datasets are reported in Section C of the supplementary file. From the results, one can see that in the case of different ratios of incomplete patterns in the training set, the performance of OAME is still better than that of other methods. This can truly reflect the fact that the OAME is superior to other methods in most datasets, especially the training and test sets with a large number of incomplete patterns.

C. Influence of \mathcal{K}_1

It worth noting that the performance of the proposed OAME method in this paper may be affected by the parameter \mathcal{K}_1 , which is the number of neighbors of the query pattern obtained

Table III: The confidence intervals of accuracy for UCI datasets by different methods with K-NN (in %).

data	ℓ	KNNI	FCMI	MF	PCC	CCAI	FID	LLA	GAIN	IVIACLR	OAME
Con	1	48.55±1.14	48.36±1.14	47.98±1.14	48.20±1.14	48.79±1.14	48.41±1.14	47.92±1.14	47.46±1.14	48.06±1.14	52.51±1.14
Con	3	45.32±1.14	45.29±1.14	45.13±1.14	45.48±1.14	45.16±1.14	44.61±1.14	45.86±1.14	42.82±1.13	45.64±1.14	47.82±1.14
Bi	5	74.43±1.18	72.61±1.20	73.94±1.18	73.56±1.19	73.98±1.18	74.62±1.17	73.86±1.19	72.27±1.21	73.67±1.19	77.84±1.12
Bi	10	73.48±1.19	71.59±1.22	73.11±1.20	73.71±1.19	73.52±1.19	73.52±1.19	73.60±1.19	70.76±1.23	72.77±1.20	82.15±1.13
Ve	7	53.81±1.50	48.27±1.51	56.97±1.49	47.71±1.51	45.96±1.50	53.85±1.50	50.69±1.51	36.12±1.45	49.50±1.51	59.53±1.48
Ve	13	49.31±1.51	37.40±1.46	49.13±1.51	32.58±1.41	45.13±1.50	51.82±1.51	42.32±1.49	25.30±1.31	32.06±1.41	55.08±1.50
Seg	8	82.16±0.70	54.44±0.91	86.10±0.63	77.97±0.76	78.28±0.75	82.49±0.69	80.66±0.72	58.44±0.90	69.92±0.84	85.82±0.64
Seg	12	76.74±0.77	34.42±0.87	77.71±0.76	69.96±0.84	72.10±0.82	78.01±0.76	73.47±0.81	38.55±0.89	34.75±0.87	82.15±1.70
Se	15	72.62±0.53	53.34±0.59	71.04±0.54	55.87±0.59	51.44±0.59	71.91±0.53	71.00±0.54	47.27±0.59	53.20±0.59	73.34±0.52
Se	18	71.72±0.53	48.24±0.59	59.60±0.58	50.22±0.59	47.43±0.59	71.06±0.54	69.66±0.55	39.40±0.58	48.32±0.59	71.74±0.53
Co	1	59.39±1.37	56.65±1.38	59.76±1.37	57.33±1.38	65.86±1.32	55.35±1.38	61.33±1.36	56.40±1.38	59.92±1.37	70.22±1.27
Co	3	53.17±1.39	45.49±1.39	52.69±1.39	48.93±1.39	60.79±1.36	49.86±1.39	54.30±1.39	41.66±1.37	51.56±1.39	61.98±1.35
Wi	5	74.20±0.86	53.74±0.98	69.16±0.91	72.26±0.88	74.22±0.86	74.08±0.86	62.04±0.95	31.96±0.91	46.68±0.98	76.04±0.84
Wi	6	55.36±0.97	39.56±0.96	35.58±0.94	54.88±0.98	56.78±0.97	54.40±0.98	40.34±0.96	25.82±0.86	35.16±0.94	58.88±0.96
Ye	4	49.37±1.30	47.68±1.30	48.63±1.30	47.50±1.30	49.59±1.30	49.15±1.30	49.01±1.30	44.86±1.29	46.13±1.30	51.34±1.30
Ye	6	42.57±1.29	42.75±1.29	43.10±1.29	43.87±1.29	39.46±1.27	42.32±1.28	39.86±1.27	40.42±1.28	43.94±1.29	44.51±1.29
Pa	8	45.10±0.94	30.40±0.87	48.05±0.95	40.52±0.93	54.80±0.94	49.40±0.95	43.63±0.94	14.54±0.67	35.81±0.91	63.81±0.91
Pa	14	31.61±0.88	15.01±0.68	27.35±0.85	26.69±0.84	42.75±0.94	34.19±0.90	28.70±0.86	7.43±0.50	16.91±0.71	48.14±0.95
Ba	/	53.56±2.30	52.78±2.31	52.11±2.31	52.89±2.31	50.22±2.31	50.67±2.31	53.67±2.30	50.11±2.31	53.11±2.31	59.11±2.27
Ave	/	58.55±1.13	47.26±1.15	56.69±1.13	53.69±1.14	56.64±1.14	58.41±1.13	55.89±1.14	41.66±1.11	48.27±1.14	64.05±1.11
WT	/	0	0	1	0	0	0	0	0	0	18

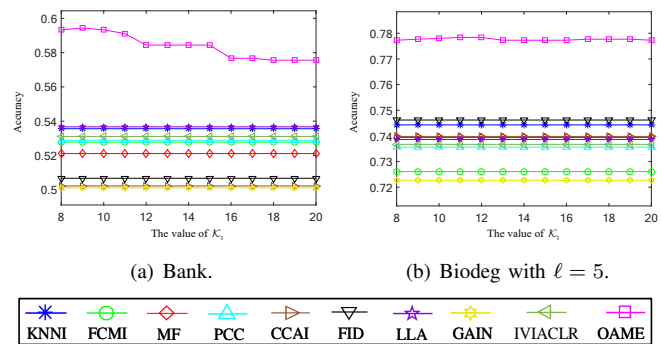
Table IV: The average classification results for Connectionist dataset with various ℓ by different methods (in %).

	KNNI	FCMI	MF	PCC	CCAI	FID	LLA	GAIN	IVIACLR	OAME
AC	41.35±1.37	32.21±1.30	37.58±1.35	36.77±1.34	46.76±1.39	39.39±1.36	39.60±1.36	30.10±1.28	34.55±1.32	48.57±1.39
PE	46.17±1.39	37.53±1.35	42.78±1.38	42.80±1.38	47.48±1.39	44.86±1.39	44.34±1.38	34.25±1.32	40.03±1.36	51.63±1.39
RE	41.99±1.37	32.32±1.30	38.08±1.35	37.07±1.35	46.87±1.39	40.08±1.37	39.93±1.36	30.30±1.28	34.78±1.33	48.84±1.39
F ₁	41.26±1.37	32.29±1.30	36.70±1.34	36.52±1.34	46.30±1.39	39.30±1.36	39.60±1.36	29.60±1.27	34.45±1.32	48.03±1.39

Table V: Classification results of Bank with various missing rates by different methods.

Method	50%				60%				70%			
	AC	PE	RE	F ₁	AC	PE	RE	F ₁	AC	PE	RE	F ₁
KNNI	53.11±2.31	54.50±2.30	53.95±2.30	51.76±2.31	53.11±2.31	55.04±2.30	54.15±2.30	50.64±2.31	53.78±2.30	56.57±2.29	54.87±2.30	50.97±2.31
FCMI	51.11±2.31	52.43±2.31	52.01±2.31	49.45±2.31	52.56±2.31	54.53±2.30	53.54±2.30	50.35±2.31	53.67±2.30	56.44±2.29	54.75±2.30	51.05±2.31
MF	51.56±2.31	52.79±2.31	52.41±2.31	50.10±2.31	52.11±2.31	53.83±2.30	53.07±2.31	49.81±2.31	53.11±2.31	55.14±2.30	54.10±2.30	50.99±2.31
PCC	52.89±2.31	54.56±2.30	53.81±2.30	51.19±2.31	53.11±2.31	55.48±2.30	54.09±2.30	50.84±2.31	52.78±2.31	55.38±2.30	53.84±2.30	50.27±2.31
CCAI	50.44±2.31	50.05±2.31	50.03±2.31	49.62±2.31	48.93±2.31	48.54±2.31	48.59±2.31	47.92±2.31	50.00±2.31	49.83±2.31	49.43±2.31	48.45±2.31
FID	51.44±2.31	52.94±2.31	52.42±2.31	49.41±2.31	51.89±2.31	53.64±2.30	52.86±2.31	49.67±2.31	54.56±2.30	56.57±2.29	55.58±2.30	52.28±2.31
LLA	51.78±2.31	53.23±2.31	52.70±2.31	50.03±2.31	50.89±2.31	51.98±2.31	51.85±2.31	48.65±2.31	53.78±2.30	56.32±2.29	54.82±2.30	51.16±2.31
GAIN	48.22±2.31	48.74±2.31	49.01±2.31	46.81±2.31	48.67±2.31	49.77±2.31	49.52±2.31	46.84±2.31	49.33±2.31	50.22±2.31	50.15±2.31	47.80±2.31
IVIACLR	52.67±2.31	54.22±2.30	53.56±2.30	51.08±2.31	53.11±2.31	55.60±2.30	54.11±2.30	50.68±2.31	52.78±2.31	54.77±2.30	53.78±2.30	50.55±2.31
OAME	57.00±2.29	57.46±2.28	57.33±2.28	56.77±2.29	58.89±2.27	59.26±2.27	59.08±2.27	58.74±2.27	58.89±2.27	59.06±2.27	58.98±2.27	58.79±2.27

by using the corresponding known attributes. The value of \mathcal{K}_1 can be slightly larger since too small \mathcal{K}_1 value cannot accurately find the most likely classes of the incomplete pattern when there are a lot of classes in dataset. Of course, it may find some noise points (isolated points) if we give too large, which will increase the risk of error. Fig. 4 clearly and intuitively reveals the influence of different \mathcal{K}_1 values in OAME based on some of the employed datasets, and the results of other datasets are shown in Section D of the supplementary file. x -coordinate corresponds to the \mathcal{K}_1 value, ranging from 8 to 20, and the y -coordinate corresponds to the average accuracy in the classification method, which is expressed in $[0,1]$. The K-NN classifier is employed to classify the test patterns, each of which contains ℓ missing values. One can see that the accuracy of different \mathcal{K}_1 values is still higher than other methods, and the variations of accuracy associated with various \mathcal{K}_1 values are very small, which confirms the robustness and effectiveness of the proposed method.

Figure 4: Classification results of different methods with various \mathcal{K}_1 values.

D. Execution time

The execution time in seconds of OAME and other comparison methods on the different datasets with K-NN as shown in Table VI. One can observe that the execution time of OAME is higher than that of KNNI and FCMI since the OAME method needs to calculate a large number of distances between patterns to obtain the neighbors in the process of imputation and the

classifier should be trained many times to optimize the training set. However, the MF and GAIN methods take a long time on building a random forest and training model, respectively. Thus, the OAME method runs considerably faster than them. In general, the proposed OAME method is more suitable for applications where high classification accuracy is required whereas efficient computation is not a strong requirement.

A dataset, named Biodeg, is employed here to investigate the impact of the parameters (*i.e.* $\mathcal{K}_1, \mathcal{K}_2, \delta, \ell$) on execution time of OAME. The EK-NN classifier is employed to classify the test patterns, and the classification result is shown in Fig.5, where x -coordinates represent the values of corresponding parameters. In particular, the x -coordinate in Fig.5(c) represents the number of δ that selected in OAME. One can see that with the increase of the values of these parameters, the execution time of OAME tends to raise with various reasons. Specially, the larger the \mathcal{K}_1 value is, the more classes may be included in neighbors, thus the more distances are needed to be computed to find neighbors from different classes in such a case. The result also reveals that a high value of \mathcal{K}_2 also brings a high execution time, since it may generate a large number of computation in the process of estimating the reliability factor in OAME. Particularly, a large number of δ will increase the time of model training during optimizing training set. Furthermore, the higher the number (*i.e.* ℓ) of missing values is, the more imprecise the classes of incomplete patterns become, which will lead to more patterns need to be filled as multiple versions based on neighbors, so the execution time is high in such a case. Although the execution time increases with increasing parameter values, it is within a reasonable range of increase, which does not affect the practical application of the proposed method. On the basis of these analyses, one can take these parameters within a reasonable range if an efficient computation is necessary for practice, especially there are a large number of missing values. In addition, inspired by CCAI [34], OAME can also take the process of optimizing a training set off-line so as to reduce the execution time.

Table VI: Execution time of different methods (In seconds).

data	KNNI	FCMI	MF	PCC	CCAI	FID	LLA	GAIN	IVI	ACL	ROAME
Ba	0.42	0.46	5.22	0.66	0.59	1.02	0.99	2.82	1.71	1.03	
Con	0.41	0.46	4.57	0.47	0.60	0.55	3.13	5.77	0.98	1.81	
Bi	0.34	0.55	24.62	0.61	0.68	0.99	1.77	6.86	2.50	2.00	
Ve	0.45	0.92	9.40	1.00	1.19	1.01	2.04	6.38	2.29	1.41	
Seg	1.06	3.03	39.86	3.92	3.27	2.28	9.36	9.54	13.8	4.11	
Se	2.33	1.59	99.12	3.09	4.70	5.18	27.89	6.39	7.69	20.69	
Co	0.46	0.79	11.41	1.26	1.59	1.09	2.29	3.61	1.94	1.81	
Wi	0.54	1.06	2.74	0.57	0.77	0.71	4.04	5.81	0.38	2.11	
Ye	0.52	0.59	4.05	0.95	0.94	0.96	2.99	7.25	1.12	1.45	
Pa	1.33	9.80	91.99	17.65	31.57	2.79	13.19	10.80	8.76	26.39	

V. CONCLUSION

To train a high-performance classifier and pursue high accuracy, we learn a credal classifier based on an optimized and adaptive multi-estimation (OAME) method for missing data imputation on training and test sets. In this work, we first introduce an optimized multi-estimation strategy to expand the training set. Then, each test pattern is adaptively estimated as one or multiple versions depending on the neighbors.

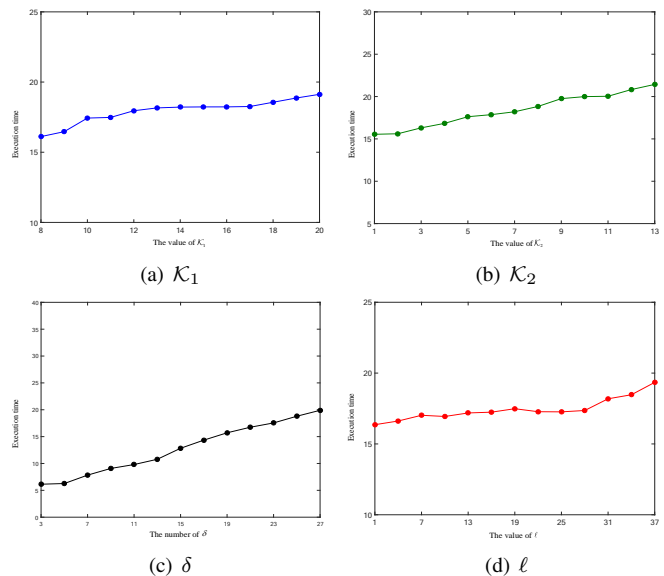


Figure 5: Execution time of OAME for Biodeg dataset with various values of parameters.

For the test pattern with multiple versions, we both estimate the importance factor of each edited version itself and the reliability factor of the classification result of the version. Finally, the corresponding outputs with different discounting factors, represented by the basic belief assignments (BBAs), are fused for credal classification based on evidence theory. Since the proposed OAME can effectively extend the training patterns while reasonably characterizing the uncertainty and imprecision caused by missing values in the imputation and classification process, its effectiveness is well verified on real datasets. In addition, we also discuss the complexity and the problems that may be encountered in applications. For example, the importance and the relevance between attributes and labels may also need to be considered for optimizing the parameter δ in applications. Over the longer-term, the goal is to develop a versatile and flexible framework for classifying missing data based on evidence theory.

REFERENCES

- [1] P. Garcia-Laencina, J. Sancho-Gomez, and A. Figueiras-Vidal, "Pattern classification with missing data: A review," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 263–282, Mar. 2010.
- [2] X. Miao, Y. Gao, S. Guo, and W. Liu, "Incomplete data management: a survey," *Frontiers of Computer Science*, vol.12, no. 1, pp. 4–25, 2018.
- [3] S. V. Buuren and C. G. M. Oudshoorn, "Multivariate imputation by chained equations: Mice v1.0 user's manual," *Technical report, TNO*, 2000.
- [4] S. V. Buuren and K. Groothuis-Oudshoorn, "MICE: Multivariate Imputation by Chained Equations in R," *Jour Syst Software*, vol. 45, no. 3, pp. 1–67, Dec. 2011.
- [5] D. J. Stekhoven and P. Bühlmann, "MissForest - nonparametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, pp. 112–118, 2012.
- [6] P. Alma, M. Ellen, C. F. Deirdre, K. Deirdre, T. M. Pham, L. Pedersen, and I. Petersen, "Missing data and multiple imputation in clinical epidemiological research," *Clin Epidemiol*, vol. 9, pp. 157–166, 2017.
- [7] J. W. Graham, "Missing Data Analysis: Making It Work in the Real World," *Annu Rev Psychol*, vol. 60, no. 1, pp. 549–576, 2009.
- [8] A. Farhangfar, L. A. Kurgan, and W. Pedrycz, "A Novel Framework for Imputation of Missing Values in Databases," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 37, no. 5, pp. 692–709, Sept. 2007.

- [9] J. Twisk, M. D. Boer, W. D. Vente, and M. Heymans, "Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis," *Jour clin epide*, vol. 66, no. 9, pp. 1022–1028, Jun. 2013.
- [10] S. Seaman, J. Galati, D. Jackson, and J. Carlin, "What Is Meant by "Missing at Random"?", *Statistical Science*, vol. 28, no. 2, pp. 257–268, May. 2013.
- [11] J. Josse and F. Husson, "missMDA: A Package for Handling Missing Values in Multivariate Data Analysis," *Jour Syst Software*, vol. 70, Apr. 2016.
- [12] Z. Wang, B. Shen, H. Shu, and G. Wei, "Quantized H_∞ Control for Nonlinear Stochastic Time-Delay Systems With Missing Measurements," *IEEE Trans. Automat. Control*, vol. 57, no. 6, pp. 1431–1444, Jun. 2012.
- [13] X. Yang, S. Yin, and O. Kaynak, "Robust Identification of LPV Time-Delay System With Randomly Missing Measurements," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 12, pp. 2198–2208, Dec. 2018.
- [14] R. J. Little and D. B. Rubin, "Statistical Analysis With Missing Data," Hoboken, NJ, USA: Wiley, 2014.
- [15] X. Miao, Y. Gao, G. Chen, B. Zheng, and H. Cui, "Processing Incomplete k-Nearest Neighbor Search," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1349–1363, Dec. 2016.
- [16] J. Dai, H. Hu, Q. Hu, W. Huang, N. Zheng, and L. Liu, "Locally Linear Approximation Approach for Incomplete Data," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1720–1732, 2018.
- [17] J. Qin, W. Fu, H. Gao, and W. X. Zheng, "Distributed k-Means Algorithm and Fuzzy c-Means Algorithm for Sensor Networks Based on Multiagent Consensus Theory," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 772–783, Mar. 2017.
- [18] N. Karmitsa, S. Taheri, A. Bagirov, and P. Makinen, "Missing Value Imputation via Clusterwise Linear Regression," *IEEE Trans. Knowl. Data Eng.*, June. 2020.
- [19] J. Liu, F. Chung, and S. Wang, "Black Hole Entropic Fuzzy Clustering," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 9, pp. 1622–1636, Sept. 2018.
- [20] S. Liu, J. Zhang, Y. Xiang, and W. Zhou, "Fuzzy-Based Information Decomposition for Incomplete and Imbalanced Data Learning," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1476–1490, Dec. 2017.
- [21] J. Yoon, J. Jordon, and M. V. D. Schaar, "GAIN: Missing data imputation using generative adversarial nets," *ICML*, pp. 5675–5684, 2018.
- [22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, and Y. S. Bengio, "Generative adversarial nets," *In Advances in Neural information processing systems*, pp. 2672–2680, 2014.
- [23] K. Jian, H. Chen, and S. Yuan, "Classification for incomplete data using classifier ensembles," in *Proc. Int. Conf. Neural Netw. Brain (ICNN&B'05), Beijing, China*, pp. 559–563, Oct. 2012.
- [24] K. Pelckmans, J. D. Brabanter, J. A. K. Suykens, and B. D. Moor, "Handling missing values in support vector machine classifiers," *Neural Netw.*, vol. 18, no. 5–6, pp. 684–692, 2005.
- [25] G. Shafer, "A Mathematical Theory of Evidence," *Princeton Univ. Press*, 1976.
- [26] P. Smets, "Decision making in the TBM: The necessity of the pignistic transformation," *Int. J. Approx. Reason.*, vol. 38, pp. 133–147, 2005.
- [27] T. Denœux, "Maximum Likelihood Estimation from Uncertain Data in the Belief Function Framework," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 119–130, Jan. 2013.
- [28] T. Denœux, "Decision-making with belief functions: A review," *Int. J. Approx. Reason.*, vol. 109, pp. 87–110, Apr. 2019.
- [29] Z. G. Liu, L. Q. Huang, K. Zhou, T. Denœux, "Combination of Transferable Classification With Multisource Domain Adaptation Based on Evidential Reasoning," *IEEE Trans. Neural. Netw. Learn. Syst.*, 2020. DOI 10.1109/TNNLS.2020.2995862.
- [30] Z. G. Liu, Y. Liu, J. Dezert, F. Cuzzolin, "Evidence combination based on credal belief redistribution for pattern classification," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 4, pp. 618–631, 2020.
- [31] A. Martin and E. Radoi, "ATR algorithms using information fusion models," *Proc. of 7th International Conference on Information Fusion (FUSION 2004), Stockholm, Sweden*, 2004.
- [32] Z. G. Liu, X. Zhang, J. W. Niu, J. Dezert, "Combination of Classifiers With Different Frames of Discernment Based on Belief Functions," *IEEE Trans. Fuzzy Syst.*, 2020. DOI 10.1109/TFUZZ.2020.2985332.
- [33] M. Vatsa, R. Singh, and A. Noore, "Unification of Evidence-Theoretic Fusion Algorithms: A Case Study in Level-2 and Level-3 Fingerprint Features," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 39, no. 1, pp. 47–56, Jan. 2009.
- [34] Z. G. Liu, Q. Pan, J. Dezert, and A. Martin, "Adaptive imputation of missing values for incomplete pattern classification," *Pattern Recognit.*, vol. 52, pp. 85–95, Apr. 2016.
- [35] Z. W. Zhang, Z. Liu, Z. Ma, Y. Zhang, H. Wang, "A new belief-based incomplete pattern unsupervised classification method," *IEEE Trans. Knowl. Data Eng.*, DOI 10.1109/TKDE.2021.3049511.
- [36] Z. G. Liu, Q. Pan, G. Mercier, and J. Dezert, "A New Incomplete Pattern Classification Method Based on Evidential Reasoning," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 635–646, Apr. 2015.
- [37] A. Martin, "Conflict management in information fusion with belief functions," *Information Quality in Information Fusion and Decision Making, Information Fusion and Data Science*, pp. 79–97, 2019.
- [38] M. E. Hellman, "The Nearest Neighbor Classification Rule with a Reject Option," *IEEE Trans. Syst. Man Cybern.*, vol. 6, no. 3, pp. 179–185, July 1970.
- [39] T. Denœux, "A k-nearest neighbor classification rule based on Dempster-Shafer Theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 5, pp. 804–813, May 1995.
- [40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20 no. 3, pp. 273–297, 1995.
- [41] A. Fitzgibbon, M. Pilu and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, 1999.
- [42] Z. G. Liu, Z. W. Zhang, Y. Liu, J. Dezert and Q. Pan, "A new pattern classification improvement method with local quality matrix based on K-NN," *Knowl.-Based Syst.*, vol. 164, pp. 336–347, 2019.
- [43] Y. M. Yang, "An evaluation of statistical approaches to text categorization," *J. Inf. Retrieval*, vol. 1, no. 1–2, pp. 67–88, 1999.
- [44] Crow, E, "Confidence Intervals for a Proportion," *Biometrika*, vol. 43, no. 3–4, pp. 423–435, 1956.



Zuo-wei Zhang was born in Anhui, China. He received the master's degree from Xi'an University of Architecture and Technology (XAUAT), China, in 2016, and is currently pursuing the Ph.D. degree at Key Laboratory of Information Fusion Technology, Ministry of Education, School of Automation, Northwestern Polytechnical University (NPU), Xi'an, China. He is also pursuing the Ph.D. degree in the laboratory IRISA, University of Rennes 1, France.

His research interest covers the theory belief functions, data mining, pattern recognition and information fusion.



Hong-peng Tian was born in Henan, China. He received the master's degree from Xi'an University of Architecture and Technology (XAUAT), China, in 2021. He is pursuing the Ph.D degree in School of Electrical Engineering, Zhengzhou University and also be a joint Ph.D student in the laboratory IRISA, Ecole Normale Supérieure de Rennes (ENS Rennes), France.

His research interest mainly focuses on the theory belief functions and its applications in pattern recognition and information fusion.



Ling-zhi Yan was born in Shaanxi, China. She received the bachelor's degree from Xi'an Aeronautical University, China, in 2018 and master's degree from Lanzhou University of Technology, China, in 2021. She works currently as a mechanical designer in Qinchuan Machine Tool&Tool Group Share Co., Ltd..

Her research interest mainly focuses on data mining, special equipment and dynamic measurement technology.



Arnaud Martin is full professor at University of Rennes 1 in the DRUID team of IRISA laboratory. He received a HDR (French ability to supervise research) in computer sciences (2009), a PhD degree in Signal Processing (2001), and Master in Probability (1998). Pr. Arnaud Martin joined the laboratory IRISA at the university of Rennes 1 as full professor in 2010 and co-create the team DRUID in 2012. He teaches data fusion, data mining, and computer sciences. He is the author of numerous papers and invited talks. He supervised numerous Phd students.

He is the founder in 2010 of the Belief Functions and Applications Society (BFAS) (www.bfasociety.org) and he is the president of the French association EGC (www.egc.asso.fr).

His research interests are mainly related to the belief functions with applications on pattern recognition, social networks and crowdsourcing.



Kuang Zhou was born in Shijiazhuang, China. He received the Bachelor degree in Chang'an University in 2010, the master degree in Northwestern Polytechnical University (NPU) in 2013, and the doctor degree in University of Rennes1 in 2016. Now he is an associate professor at School of Mathematics and Statistics in NPU.

His research work focuses on the theory of belief functions and its applications.