# A New Belief-based Incomplete Pattern Unsupervised Classification Method

Zuowei Zhang, Zhe Liu, Zongfang Ma, Yiru Zhang, Hao Wang

# A New Belief-based Incomplete Pattern Unsupervised Classification Method

Zuo-wei Zhang[a], Zhe Liu[b], Zong-fang Ma[b], Yi-ru Zhang[a], Hao Wang[c]
$a$. Institut de Recherche en Informatique et Systèmes Aléatoires, University of Rennes 1, France.
$b$. School of Information and Control Engineering, Xi'an University of Architecture and Technology, China.
$c$. Department of Computer Science, Norwegian University of Science and Technology, Norway.

***Abstract*—The clustering of incomplete patterns is a very challenging task because the estimations may negatively affect the distribution of real centers and thus cause uncertainty and imprecision in the results. To address this problem, a new belief-based incomplete pattern unsupervised classification method (BPC) is proposed in this paper. Firstly, the complete patterns are grouped into a few clusters by a classical soft method like fuzzy $c$-means to obtain the corresponding reliable centers and thereby are partitioned into reliable patterns and unreliable ones by an optimization method. Secondly, a basic classifier trained by reliable patterns is employed to classifies unreliable patterns and the incomplete patterns edited by the neighbors. In this way, most of the edited incomplete patterns can be submitted to specific clusters. Finally, some ambiguous patterns will be carefully repartitioned again by a new distance-based rule depending on the obtained reliable centers and belief functions theory. By doing this, a few patterns that are very difficult to classify between different specific clusters will be reasonably submitted to meta-cluster which can characterize the uncertainty and imprecision of the clusters due to missing values. The simulation results show that the BPC has the potential to deal with real datasets.**

**Keywords:** clustering, incomplete pattern, belief functions theory, unsupervised classification, fuzzy $c$-means.

## I. INTRODUCTION

Cluster analysis [1]-[4], also known as unsupervised classification, is an important branch of pattern recognition and machine learning, which can classify massive data without any prior information into the same groups with similar data structures or physical structures and has extensive applications in many fields, such as financial analysis, medical diagnosis, image processing, information retrieval and bioinformatics. A large number of clustering methods [4]-[11] have been developed and they can be divided into two types, namely hard clustering [4], [5] and soft clustering [6], [11]. Among them, the $K$-means algorithm [5] is one of the most well-known hard clustering techniques which allows each pattern to belong to only one cluster, and the fuzzy $c$-means [6] is the most widely used soft clustering method where each pattern can be committed to different clusters with different probabilities. These classical unsupervised classification methods are generally applicable to complete data analysis, whereas the pattern with missing values is a common issue in many real-world datasets. For example, $45\%$ of the datasets in the UCI [12], as one of the standard repository commonly used in machine learning algorithms, contains missing values caused by many manifold reasons, including respondents refuse to answer some privacy questions when filling out a questionnaire and patient data or treatment data can not be obtained completely because of confidentiality in medical statistics, and so on. Therefore, preprocessing is the prerequisite for the use of these methods.

A number of methods [13] have emerged for preprocessing missing data depending on three types of missing data mechanisms summarized by Little and Rubin [14] according to the cause of the missing. The simplest whole data strategy also called the discarding method, is to ignore all incomplete patterns if they only take a small amount of the entire dataset, and then cluster the complete patterns. In many cases, the missing values are usually estimated by imputation strategies [15]-[24], and then the incomplete patterns with the estimations are clustered. For instance, one of the simplest and fast mean imputation (MI) method is introduced in [15]. The missing values are replaced by the average values of the same attributes of all complete patterns. A commonly used weighted $K$-nearest neighbor imputation (KNNI) method is introduced in [18] based on DNA microarray data to reasonably simulate the different effects of different neighbors on incomplete patterns. In fuzzy $C$-means imputation (FCMI) method [19], [20], the missing values of each incomplete pattern are estimated based on the membership degree and clustering center generated by FCM [6]. In POCS [21], according to the KNNs distribution density of missing values, the probabilistic information granules with missing values are introduced into the optimal whole strategy (OCS) of incomplete pattern clustering to obtain classification results. Particularly, a new method, named fuzzy-based information decomposition (FID) [22], is developed to simultaneously address class imbalance and missing values, and they are treated as the same missing data estimation problem. Recently, a new linear local approximation (LLA) [23], uses the optimal weights of KNNs obtained by local linear reconstruction to estimate the missing values. Interestingly, some recent works are dedicated to multiple estimations or non-estimation of missing values [26], [27]-[28]. These methods have achieved good results. However, these methods may lead to the following problems:

(1) Estimation deviations may be large or too much useful information lost, which will have a negative impact on the performance of clustering methods. For example, in the discarding method [13], some useful information may be lost, which is difficult to obtain in reality and will cause great waste. The correlation between data is not taken into account in MI [15], so the estimations may be quite different from the truth.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



(2) Most clustering algorithms tend to directly cluster the edited data with estimations, which ignore the negative impact of imputation strategies on the complete pattern. For instance, if there is a large difference between the estimations and the truth, the estimations may change the initial distribution of the original data, and then may change the initial position of the cluster center while having a negative impact on the clustering of complete patterns, which is very likely to happen.

(3) They do not reasonably characterize the uncertainty and imprecision caused by missing values or the lack of information on the overlapping of different clusters. The estimations obtained by different imputation strategies are often different, which indicates that the estimations cannot replace the role of the real values and inevitably lead to uncertainty. At the same time, some patterns in the overlapping areas of different clusters may be difficult to be clustered to a specific cluster. If they are forced to cluster to a singleton cluster, the risk of error will be greatly increased.

In this paper, we propose a new belief-based unsupervised classification method (BPC) for dealing with incomplete patterns. The core is to adopt a more cautious strategy to characterize the uncertainty and imprecision caused by missing values or data distribution under belief functions theory.

Belief functions theory [29], [30] has been applied in many fields, such as data clustering [31]-[34], data classification [35]-[37] and decision-making [38]-[39]. In belief functions, the pattern is not only allowed to belong to any singleton (specific) cluster, but also to different meta-clusters (denoted as particular disjunctions of several singleton clusters) with different possibilities which can well represent the imprecision of clustering for uncertain patterns. For instance, the EVCLUS algorithm [31] is introduced to cluster the relational data with the large dissimilarities between objects, and its improved version for computational speed is developed in [32].

In ECM, it produce three kinds of clusters with different masses of beliefs. They are singleton clusters (e.g. $\omega_i$), meta-clusters (e.g. $\omega_j \cup ... \cup \omega_k$) defined by disjunction (union) of several singleton clusters, and the outlier cluster represented by $\emptyset$. The meta-cluster center is considered to be the average of the involved singleton cluster centers and the probability of pattern belonging to the meta-cluster only depends on the distance between the pattern and the center, which is unreasonable because it may cause clustering error when the center of a meta-cluster is very close to the center of a singleton cluster. The limitation is overcome in [11] where the pattern committed to a meta-cluster is not only close to the center of the meta-cluster but also close to the related singleton clusters included in the meta-cluster. However, these clustering methods are mainly developed to complete data clustering and they do not consider the situation of incomplete data. Recently, some researches are devoted to the classification of incomplete data, which effectively complements the application fields of belief functions theory. For instance, a new incomplete pattern belief classification [36] method, where the incomplete pattern is directly assigned into a specific class or filled by KNNs in terms of the class information of its neighbors, and then, a weighted possibility distance method is designed to partition the uncertain incomplete pattern under the framework of belief functions. Whereas there is no relevant literature(s) on the clustering of incomplete data.

The main contribution of this paper is the development of a new efficient belief-based unsupervised classification method that works on belief function theory for incomplete patterns (BPC). It focuses on avoiding cluster center shifts caused by estimations and the uncertainty and imprecision due to missing values. Specifically, BPC first adopts the existing classical soft clustering methods (such as FCM [6], NC [9]) to cluster the complete patterns[1], and thereby adaptive partitions them as reliable complete patterns or unreliable ones based on step iteration. Then, a basic classifier such as K-NN [40], SVM [41], NB [42] trained by the reliable patterns is employed to classifies unreliable patterns and the edited incomplete patterns estimated by the neighbors. In this way, most of the edited incomplete patterns can be assigned to specific clusters. Whereas if the possibilities of a pattern being classified into different singleton clusters are not distinct different, we consider it to be an uncertain pattern[2]. Finally, these uncertain patterns will be carefully repartitioned again by a new distance-based rule based on belief functions theory. By doing this, a few patterns that are difficult to classify in specific clusters will be reasonably assigned to meta-clusters which can characterize the uncertainty and imprecision due to missing values.

In BPC, meta-clusters will be conditionally kept for the uncertain patterns that are hard to correctly classify, which can not only reduce the error rate but also reasonably characterize the uncertainty and inaccuracy caused by missing values or the lack of information on the overlapping of different clusters. Once the pattern is submitted to meta-cluster, it indicates that the singleton clusters contained in the meta-cluster are indistinguishable for the pattern, and the existing information cannot accurately partition the pattern. The introduction of meta-clusters can effectively reduce the misclassification rate. In many practical problems, we prefer to obtain some imprecise results that need to be judged by other methods, rather than take a greater risk to get a wrong classification. Moreover, sometimes more precise but expensive methods can be used to make auxiliary judgments for difficult patterns. However, if these precise technologies are used at the very start, they will be expensive. Therefore, we can assign the hard-to-distinguish patterns to meta-clusters, and then confirm the patterns in meta-clusters only further at a much lower cost.

This paper is organized as follows. Section II briefly introduces the background knowledge. Section III introduces the new BPC method. Then in Section IV, the dataset of UCI repository is used to test the proposed BPC method and compare it with other methods for dealing with incomplete data. The conclusion of this paper is given in Section V.

[1]Here we will choose FCM to cluster complete patterns, because it is the most classical representative of soft clustering algorithm which has mature theory and is widely used.

[2]Once a pattern is regarded as an uncertain pattern, it indicates that the pattern may originally be distributed in overlapping regions of different clusters or caused to be distributed in overlapping regions of different clusters due to missing values.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



## II. BACKGROUND

### A. Fuzzy c-means algorithm

As one of the most famous soft clustering methods, FCM has been widely used in various fields [6], [7] and its core idea is that the patterns belonging to one cluster have a high degree of membership, while those belonging to the others have a relatively low degree of membership. It can divide a complete dataset $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^p$ into $c$ clusters characterized by prototypes $V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c\} \in \mathbb{R}^{c\times p}$ depending on minimizing the following objective functions:

$$J_{FCM}(U,V) = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^{\beta} d_{ij}^2, \tag{1}$$

with

$$\sum_{k=1}^{c} u_{ik} = 1, \quad \text{for} \quad i = 1,2,\ldots,N, \tag{2}$$

where $\beta \in (1,\infty)$ denotes the fuzzification parameter, which will determine the fuzziness degree of clustering analysis results; $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]^T$ is a pattern datum, and $x_{ik}$ is the $k$-th attribute value of $\boldsymbol{x}_i$; $\boldsymbol{v}_j$ is the $j$-th cluster prototype, $\boldsymbol{v}_j \in \mathbb{R}^p$; $u_{ij}$ is the degree of membership of $\boldsymbol{x}_i$ in the $j$-th cluster, and let the partition matrix $U = [u_{ij}] \in \mathbb{R}^{n\times c}$; $d_{ij}$ denotes the Euclidean distance between $\boldsymbol{x}_i$ and the cluster center $\boldsymbol{v}_j$.

The necessary conditions for minimizing (1) with the constraint (2) result in the following update formulas:

$$\boldsymbol{v}_j = \frac{\sum_{i=1}^{N} u_{ij}^{\beta}\boldsymbol{x}_i}{\sum_{i=1}^{N} u_{ij}^{\beta}}, \quad \text{for} \quad j = 1,2,\ldots,c, \tag{3}$$

and

$$u_{ij} = \frac{d_{ij}^{-2/(\beta-1)}}{\sum_{k=1}^{c} d_{ik}^{-2/(\beta-1)}}, \quad \text{for} \quad i = 1,2,\ldots,n \quad \text{and} \quad j = 1,2,\ldots,c, \tag{4}$$

The procedure of FCM is to optimize the clustering objective function (1) by alternating optimization, that is, the minimization steps (3) and (4) are repeated until the change of the objective function below a certain threshold.

### B. $K$-nearest neighbors imputation

Since weight $K$-nearest neighbor imputation (WKNNI) [18] is employed in this paper to estimate the missing values, we briefly recall the principle of KNNI here. Let us consider a complete dataset $X_{com}$ with $p$-dimensional space in the class editing framework $\Omega = \{\omega_1, ..., \omega_c\}$ and a query pattern $\boldsymbol{x}_i$ with $t$ $(1 \leq t < p)$ known attributes. For estimating the missing values, one can obtain the neighbors of $\boldsymbol{x}_i$ first by calculating the distances based on the known attributes. The Euclidean distance $d_{ij}$ between the pattern $\boldsymbol{x}_i$ and each pattern $\boldsymbol{x}_j$ included in the set $X_{com}$ is defined by:

$$d_{ij} = \sqrt{\sum_{l=1}^{t}(x_{il} - x_{jl})^2}, \tag{5}$$

Then the $K$ minimum distances $d_{ik}$ and corresponding complete patterns $\boldsymbol{x}_k \in X_{com}$, $k = 1,2,\ldots,K$, are obtained. Noted that the distance between the pattern $\boldsymbol{x}_i$ and KNNs may vary when estimating missing values, thus the weights are reasonably applied to weigh differently the impact of KNNs in obtaining the estimations. In WKNNI, the weight $\alpha_i^k$ of the $k$-th neighbor $\boldsymbol{x}_k$ is denoted as:

$$\alpha_i^k = \frac{e^{-d_{ik}}}{\sum_{k=1}^{K} e^{-d_{ik}}} \tag{6}$$

Once the KNNs of pattern $\boldsymbol{x}_i$ and weight $\alpha_i^k$ are obtained, the missing value $\widetilde{x}_{ij}$ of the pattern $\boldsymbol{x}_i$ is thereby estimated in the same dimension as follows:

$$\widetilde{x}_{ij} = \sum_{k=1}^{K} \alpha_i^k \cdot x_{kj}, \tag{7}$$

By doing this, one can deal with incomplete patterns with estimations in the same way as to complete patterns.

### C. Belief functions theory

Belief functions theory [29], [30] was first proposed by Dempster and formed by Shafer generalization, which is also known as Evidential Reasoning or Dempster-Shafer theory (DST). It is a theoretical framework for reasoning with partial and unreliable information, which can deal with uncertain problems well. In this theory, a set of finite mutually exclusive and complete elements $\Omega = \{\omega_1, \omega_2, \cdots, \omega_c\}$ is defined as the framework of discernment of the problem under study. The set of all subsets of $\Omega$ is called the power-set of $\Omega$, which is represented as $2^\Omega$ and contains $2^{|\Omega|}$ elements. For example, if $\Omega = \{\omega_1, \omega_2, \omega_3\}$, then $2^\Omega = \{\emptyset, \omega_1, \omega_2, \omega_3, \omega_1 \cup \omega_2, \omega_1 \cup \omega_3, \omega_2 \cup \omega_3, \Omega\}$. The singleton cluster (e.g. $\omega_i$) is also called specific cluster. The disjunction of several singleton elements (e.g. $\omega_i \cup \omega_j, \omega_i \cup \omega_j \cup \omega_k$) represents the partial ignorance in $2^\Omega$, and they are called meta-clusters.

The basic belief assignment (BBA) on the framework of discernment $\Omega$ is a function $m(.)$ from $2^\Omega$ to $[0,1]$, and satisfies the following conditions:

$$\begin{cases} \sum_{A\in 2^\Omega} m(A) = 1 \\ m(\emptyset) = 0 \end{cases} \tag{8}$$

All the elements $A \in 2^\Omega$ such that $m(A) > 0$ are called the focal elements of $m(.)$. A credal partition [10], [31] is defined as the $N$-tuple $M = (m_1, \ldots, m_N)$, where $m_i$ is the BBA of the pattern $\boldsymbol{x}_i \in X$, $i = 1,2,\ldots,N$, associated with the different elements of the power-set $2^\Omega$.

Some evidential clustering methods [10], [11] have been developed to address the uncertain and imprecise problem in partitioning the unlabeled dataset. Here we briefly recall the representative evidential $c$-means (ECM) [10] proposed by Denœux, which is considered to be a very important component in the clustering task. It can be regarded as an evidential version of the fuzzy $c$-means algorithm (FCM [6] and noise clustering algorithm (NC) [9]. In ECM, the pattern is assigned not only to singleton clusters including the outlier cluster, but also to any meta-clusters of $2^\Omega$ by a "mass of belief" which is similar to the membership degree in FCM and denoted as $m(.)$. The objective function $J_{ECM}$ can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



written as follows:

$$J_{ECM}(M,V)=\sum_{i=1}^{N}\sum_{\{j/A_j\neq\emptyset,A_j\subseteq\Omega\}}|A_j|^{\alpha}m_{ij}^{\beta}d_{ij}^{2}+\sum_{i=1}^{N}\delta^2 m_{i\emptyset}^{\beta} \tag{9}$$

subject to

$$\sum_{\{j/A_j\neq\emptyset,A_j\subseteq\Omega\}} m_{ij}+m_{i\emptyset}=1,\ i=1,...,N \tag{10}$$

where $\alpha$ is an adjustable weight parameter of the distance between the pattern and the cluster center, and $|A_j|$ denotes the cardinality of $A_j$. If $A_j=\{\omega_k,\omega_t\}$, then $|A_j|=2$.

In ECM, the mass of the belief $m_{ij}$ of the pattern $\boldsymbol{x}_i$ can be obtained by minimizing the objective function (9) with constraint (10), and it is given by:

$$m_{ij}=\frac{|A_j|^{-\alpha/(\beta-1)}d_{ij}^{-2/(\beta-1)}}{\sum\limits_{A_k\neq\emptyset}|A_j|^{-\alpha/(\beta-1)}d_{ij}^{-2/(\beta-1)}+\delta^{-2/(\beta-1)}}. \tag{11}$$

In applications, ECM may produce unreasonable results when some singleton cluster centers are close to that of meta-clusters. Since the definition of the distance between patterns and meta-cluster centers is not precise and it has been improved by a new method, called credal $c$-means (CCM) [11]. Inspired by these methods, we design a new distance-based rule for classifying uncertain patterns to characterize the imprecision caused by missing values in this paper.

## III. BELIEF-BASED UNSUPERVISED CLASSIFICATION METHOD

A new BPC method is proposed to deal with incomplete patterns based on belief functions theory. BPC firstly clusters the complete patterns by the famous FCM method and thereby obtain the membership of the pattern for different clusters. Some of the patterns with clear cluster information are naturally partitioned into reliable patterns, which are regarded as having strong support for a specific cluster. By contrast, the remaining complete patterns without support a particular cluster are partitioned as unreliable patterns, some of which may be in overlapping regions of different clusters. Then, these reliable patterns are used to train a basic classifier and thus to classify unreliable patterns and the edited incomplete patterns estimated by the neighbors. For patterns with strong support for specific clusters in the classification results, they are partitioned into singleton clusters. By contrast, for those uncertain patterns that do not clearly belong to singleton clusters, a new and more cautious distance-based rule method is designed to submit some of them to the meta-clusters to reduce the risk of misclassification. The classification of the uncertain pattern in meta-cluster can be eventually precise using some other (costly) techniques or with extra information sources. By doing this, the BPC method can prevent us to take erroneous fatal decisions whenever it is essential to do it.

### A. Optimization of reliable patterns

Let us consider a dataset $X$ in $p$-dimensional real space to be clustered into $c$ specific clusters under the frame of discernment $\Omega=\{\omega_1,\ldots,\omega_c\}$. The set $X$ can be divided into two parts: the set $X_{com}$ contains complete patterns and the set $X_{mis}$ contains incomplete patterns, i.e., $X=X_{com}+X_{mis}$. Firstly, FCM[3] is employed to cluster complete patterns and thereby to obtain different memberships $m_{ij}$ of each pattern $\boldsymbol{x}_i\in X_{com}$ belonging to the different specific clusters (such as $\omega_j$). In this way, the complete patterns can be divided into reliable and unreliable ones by a threshold $\gamma$, and denoted as:

$$X_{re}=\{\boldsymbol{x}_i|m_i^{\max}(\omega_{\max})\geq\gamma\} \tag{12}$$

with

$$m_i^{\max}(\omega_{\max})=\max\{m_i(\omega_1),\ldots,m_i(\omega_c)\} \tag{13}$$

where $m_i(\omega_j)$ is the membership of $\boldsymbol{x}_i$ belonging to the cluster $\omega_j$ ($j=1,\ldots,c$). $m_i^{\max}$ ($\omega_{\max}$) denotes the support (membership) degree of $\boldsymbol{x}_i$ belonging to the most believed cluster $\omega_{\max}$ ($\omega_{\max}\in\Omega$). $X_{re}$ is the set of all the complete patterns satisfying condition $m_i^{\max}(\omega_{\max})\geq\gamma$, and $\gamma$ is the threshold to control the number of patterns in $X_{re}$.

One can find that all the patterns in $X_{re}$ have distinct class information with the increase of $\gamma$, especially when $\gamma$ approaches 1. This indicates that the pattern $\boldsymbol{x}_i$ is very close to the center of the cluster $\omega_{\max}$, and $\boldsymbol{x}_i$ can be considered as a reliable pattern belonging to $\omega_{\max}$. Conversely, the remaining patterns with satisfying Eq. (12) in $X_{com}$ are regarded as unreliable ones and therefore form the unreliable pattern set $X_{unr}$, $X_{com}=X_{re}+X_{unr}$. The pattern in $X_{unr}$ is considered to be close to the centers of multiple clusters at the same time caused by a variety of reasons. For example, the membership values are very close if the pattern is in the overlapping or middle region of different clusters.

Since $\gamma$ determines the number of reliable patterns of different clusters and these reliable patterns then are used to train the basic classifier, the value of $\gamma\in[\gamma_{Bel},\gamma_{Pl}]$ is a key parameter in BPC, where $\gamma_{Bel}$ is the belief value of $\gamma$ (i.e., the lowest value that $\gamma$ can take) and $\gamma_{Pl}$ is the plausibility value of $\gamma$ (i.e., the desirable maximum of $\gamma$). If $\gamma_{Bel}$ is too small, a number of unreliable patterns will be partitioned into $X_{re}$, which is not a good choice for training the basic classifier; By contrast, too big $\gamma_{Pl}$ will lead to an increase in the number of patterns in the set $X_{unr}$ while that in the set $X_{re}$ decrease. In such a case, it will also have a negative impact on the performance of the basic classifier if the reliable patterns of different clusters are insufficient[4].

Here, we provide a new adaptive method for obtaining more accurate $\gamma$ values for a particular dataset $X$. At its core, the basic classifier achieves its best performance when the distance between the probability result of clustering patterns into different clusters in $X_{com}$ and the probability result of the classifier classifying patterns into different classes is smallest. For a particular $\gamma_s$, we define the probabilistic distance $\Psi(\gamma_s)$

[3]Here we assume that the relevant methods (e.g., FCM [6], K-NN [40]) introduced by default when applying the proposed method are also applicable.

[4]After many tests conducted on various real datasets, this method generally produces good performances with $\gamma\in[0.6,0.95]$ in practice, i.e., $\gamma_{Bel}=0.6$ and $\gamma_{Pl}=0.95$, and we recommend $\gamma=0.85$ as the default.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



obtained for the complete patterns as follows:

$$\Psi(\gamma_s) = \sum_{i=1}^{N} \sum_{j=1}^{c} \|p_{ij} - m_{ij}\|^2 \tag{14}$$

where $m_{ij}$ is the membership of $\boldsymbol{x}_i$ in $X_{com}$ belonging to $\omega_j$ $(j = 1, \ldots, c)$ obtained from FCM; $n$ is the number of the reliable patterns in $X_{com}$ by $\gamma_s$, and $c$ is the real cluster number; $\|\cdot\|^2$ represents the Euclidean distance and $p_{ij}$ is the probability of $\boldsymbol{x}_i$ belonging to class $\omega_j$ obtained by $\Gamma(.)$, denoted as:

$$\boldsymbol{p}_i = \Gamma(\boldsymbol{x}_i \,|X_{re}^s) \tag{15}$$

where $\boldsymbol{p}_i$ is the probability of $\boldsymbol{x}_i$ being assigned to different classes using the basic classifier $\Gamma(.)$ trained by the set $X_{re}^s$ and $\boldsymbol{p}_i = [p_i(\omega_1), \ldots, p_i(\omega_c)]$ with $p_{ij} \triangleq p_i(\omega_j)$. The choice of the classifier $\Gamma(.)$ is left to one's preference. For instance, one can use the K-NN [40], SVM [41] or NB [42], etc. Here we choose the K-NN as the basic classifier in this paper.

One can easily find that with $\gamma_s$ takes different values, $\Psi(\gamma_s)$ will also get different corresponding values. In other words, $\gamma_s \in [\gamma_{Bel}, \gamma_{Pl}]$ is the optimal value for the specific dataset $X$ if $\Psi(\gamma_s)$ reaches the minimum, denoted as:

$$\Psi(\gamma_{opt}) = \min\{\Psi(\gamma_1), \ldots, \Psi(\gamma_{\mathcal{S}})\} \tag{16}$$

where $\Psi(\gamma_{opt})$ is the probabilistic distance for the optimal solution $\gamma_{opt}$ and $\gamma_{opt} \in [\gamma_1, \ldots, \gamma_s, \ldots, \gamma_{\mathcal{S}}]$[5]; $\mathcal{S}$ is the total number of iterations and determined as follows:

$$\mathcal{S} = \frac{\gamma_{Pl} - \gamma_{Bel}}{\lambda} + 1 \tag{17}$$

where $\lambda$ is the step size of each iteration. Thus, the value of $\gamma_s$ (if iterated to the $s$-th) is:

$$\gamma_s = \gamma_{Bel} + \lambda(s-1), s = 1, \ldots, \mathcal{S} \tag{18}$$

1) **Guideline for Choosing Parameter** $\lambda$: The selection of $\lambda$ plays an important role in the generation of the optimal threshold. If $\lambda$ is too big, we may not be able to get the optimal $\gamma_{opt}$ (or near the optimal) and then have a negative impact on the performance of the classifier. Too small $\lambda$, however, will lead to increased computational complexity in the process of obtaining the optimal $\gamma_{opt}$. According to many exploratory experiments, $\lambda = 0.01$ can be chosen as the default.

***Discussion 1***: FCM is used as the basic unsupervised method for clustering the complete patterns in $X_{com}$ in this paper. In applications, many classical soft clustering algorithms can be used for clustering. In traditional probability frameworks, such as FCM [6], PCM [8] and NC [9], bayesian probability is used to denote the possibility that patterns belong to different singleton (specific) clusters. Nevertheless, the framework of pattern classification is extended to the power-set $2^\Omega$ in belief functions theory, which allows the pattern to have different BBAs for not only the different singleton (specific) clusters but also the meta-clusters (such as ECM [10] and CCM [11]). In this case, BPC only considers the BBAs that the patterns are clustered into singleton clusters and regards that the patterns clustered into meta-clusters are unreliable and they are submitted directly to the set $X_{unr}$.

***Discussion 2***: In BPC, the optimal $\gamma_{opt}$ is obtained by equivalent step $\lambda$ iteration, which may cause a lot of computational cost. In applications, one can obtain $\gamma_{opt}$ by optimization method. Specifically, a number of points are randomly selected between $\gamma_{Bel}$ and $\gamma_{Pl}$, and the corresponding $\Psi(\gamma_s)$ values can be calculated by Eq. (14). Then, the corresponding relationships between $\gamma_s$ and $\Psi(\gamma_s)$ are used for curve fitting and the fitting function $\Psi(\gamma_s) = f(\gamma_s)$ can be obtained through a variety of methods, such as MATLAB curve fitting experiment. Finally, the optimal $\gamma_{opt}$ can be gotten by finding the extremum (minimum) point of the fitting function $f(\gamma_s)$.

### B. Decision-making of remaining patterns

Since we are concerned with uncertainty and imprecision due to missing values rather than imputation methods, different estimation strategies can in principle be chosen as potential alternatives, and here we use one of the simplest and most commonly used KNNI method [18]. In other words, missing values of incomplete patterns are naturally estimated by the neighbors with different weights defined in Eqs. (5)-(7). These edited patterns with estimations and the unreliable patterns thereby can be classified by the trained classifier $\Gamma(\cdot)$ defined in Eq. (15). By doing this, one can obtain the probability of these patterns belonging to different clusters. Generally, the larger the value $p_{ij}$, the greater the probability that the pattern $\boldsymbol{x}_i$ belongs to cluster $\omega_j$. In this way, most of these patterns can be submitted to specific clusters. However it may also happen that the probability values of the pattern being allocated to different clusters are very close, which means that the pattern, called uncertain pattern, may be in the overlapping areas of different clusters or the estimations of missing values of the pattern are not accurate enough. If forced to be classified into a singleton cluster, it will be result a greater risk of misclassification. Therefore, a more cautious method is needed to repartition these uncertain patterns to characterize the uncertainty and imprecision caused by the lack of information or missing values.

Based on the above analysis, we adopt a flexible and adjustable method based on the belief functions theory to repartition uncertain patterns, which is an appropriate compromise between the error rate and the imprecision rate by adjusting parameter. For the uncertain pattern $\boldsymbol{x}_i$ belonging to some different singleton (specific) clusters (e.g., $\omega_t$ and $\omega_k$) with little difference in probability, we assume that there exists one cluster, called the optimal meta-cluster $\Lambda_i$ (i.e., $\Lambda_i = \omega_t \cup \omega_k$) for $\boldsymbol{x}_i$, composed of those singleton clusters (i.e., $\omega_t$ and $\omega_k$). The uncertain pattern $\boldsymbol{x}_i$ then will be repartitioned again under the editing framework $\Omega^i = \{\omega_1, ..., \omega_c, \Lambda_i\}$ after finding the optimal meta-cluster $\Lambda_i$.

For simplicity and notation convenience, we assume that the most likely cluster is $\omega_{\max}$, that is $p_i(\omega_{\max}) = \max\{p_{ij}\}$, $j = 1, 2, \ldots, c$. In the classification result, the cluster $\omega_{\max}$ with the biggest probability is denoted as:

$$p_i(\omega_{\max}) = \max\{p_i(\omega_1), \cdots, p_i(\omega_c)\}, \tag{19}$$

[5]In this case, the $\gamma_{opt}$ we get may not be the optimal value in the mathematical sense, but since the step size $\lambda$ takes a small value, i.e.,$\lambda = 0.01$, $\gamma_{opt}$ can be approximated as the optimal value in practical applications. In addition, we also give an alternative in ***Discussion 2***.

In practice, $p_i(\omega_j)$ may be very close to $p_i(\omega_{\max})$, while $\omega_j$ and $\omega_{\max}$ are different clusters. In such case, the pattern can also potentially belong to $\omega_j$, and we should consider all possible singleton clusters as potential solution. The set of these clusters is considered as $\Lambda_i$ i.e., the optimal meta-cluster $\Lambda_i$, and it is defined by:

$$\Lambda_i = \{\omega_{max} \cup, ..., \cup \omega_j | p_i(\omega_{max}) - p_i(\omega_j) < \epsilon\}, \quad (20)$$

where $\epsilon$ has only a small adjustable threshold. The singleton clusters in $\Lambda_i$ are indistinguishable for the pattern $\boldsymbol{x}_i$ under the threshold $\epsilon$. Inspired by [10], [11], we can repartition the pattern $\boldsymbol{x}_i$ by a distance-based rule to reduce the risk of misclassification once obtaining the optimal meta-cluster $\Lambda_i$.

In traditional evidential clustering [10], [11], the belief degree for one pattern belonging to different clusters (including singleton cluster and meta-cluster) depends mainly on the distance between the pattern and the cluster centers. The smaller the distance between the pattern and the clustering center, the higher the degree of belief that the pattern belongs to the corresponding cluster. It is easy to see that the essence of the degree of belief is the inverse distance relationship between the pattern and the clustering center. Therefore, we provide a distance-based rule here to classify the uncertain patterns directly. The core is that the pattern $\boldsymbol{x}_i$ will be submitted to the cluster if the distance between the pattern and the cluster center (including all singleton clusters and the optimal meta-cluster $\Lambda_i$) is the smallest. Since the reliable clustering centers have been obtained based on FCM, we therefore define the Euclidean distance between the pattern and the singleton cluster cluster as the distance, which is similar as ECM [10] and CCM [11]. It should be noted that the distance between the pattern and the optimal meta-cluster $\Lambda_i$ is depending not only on the Euclidean distance between the pattern and the optimal meta-cluster center (i.e. the mean value of the involved singleton cluster centers) but also on the Euclidean distance between the pattern and all the singleton cluster centers involved in the meta-cluster. The distance $\overline{d}_{(\boldsymbol{x}_i,\Lambda_i)}$ between the pattern $\boldsymbol{x}_i$ and the optimal meta-cluster $\Lambda_i$ is calculated as follows:

$$\overline{d}_{(\boldsymbol{x}_i,\Lambda_i)} = \frac{\sum\limits_{\omega_k \in \Lambda_i} d_{ik} + d^*_{(\boldsymbol{x}_i,\overline{\boldsymbol{v}}_{\Lambda_i})}}{|\Lambda_i| + 1}, \quad (21)$$

where $\overline{\boldsymbol{v}}_{\Lambda_i} = \frac{\sum_{\omega_k \in \Lambda_i, |\omega_k|=1} \boldsymbol{v}_k}{|\Lambda_i|}$, and $\boldsymbol{v}_k$ is the center of the singleton cluster $\omega_k$ involved in $\Lambda_i$; $d^*_{(\boldsymbol{x}_i,\overline{\boldsymbol{v}}_{\Lambda_i})} = ||\boldsymbol{x}_i - \overline{\boldsymbol{v}}_{\Lambda_i}||^2$ is the distance between $\boldsymbol{x}_i$ and the center $\overline{\boldsymbol{v}}_{\Lambda_i}$ of the optimal meta-cluster $\Lambda_i$; $|\Lambda_i|$ denotes the cardinality of $\Lambda_i$, that is, the number of singleton clusters it contains.

After obtaining the distance between the uncertain pattern $\boldsymbol{x}_i$ and the clusters included in the editing framework $\Omega_i$, one can easily partition $\boldsymbol{x}_i$ into one singleton cluster (e.g., $\omega_k$) or the optimal meta-cluster $\Lambda_i$. If the distance $\overline{d}_{(\boldsymbol{x}_i,\Lambda_i)}$ is less than the distance from $\boldsymbol{x}_i$ to each singleton cluster center involved in the meta-cluster $\Lambda_i$, BPC will partition $\boldsymbol{x}_i$ into the optimal meta-cluster $\Lambda_i$, denoted as:

$$\Lambda_i = \{\boldsymbol{x}_i | \overline{d}_{(\boldsymbol{x}_i,\Lambda_i)} < min\{d_{ik}\}\}, \quad \omega_k \in \Lambda_i. \quad (22)$$

It is important to note that although each uncertain pattern has its corresponding optimal meta-cluster, it does not mean that the pattern will be partitioned into the meta-cluster, since the choice of the optimal meta-cluster is highly dependent on the parameter $\epsilon$. For uncertain pattern $\boldsymbol{x}_i$, $\boldsymbol{x}_i$ will naturally be submitted to the corresponding optimal meta-cluster $\Lambda_i$ if the value of $\epsilon$ is reasonable. By contrast, the pattern $\boldsymbol{x}_i$ will still be submitted to the singleton cluster that $\boldsymbol{x}_i$ most likely to belong to. The following is the guidance on the parameter $\epsilon$.

2) **Guideline for Choosing Meta-cluster Threshold** $\epsilon$: In applications, the threshold of meta-cluster $\epsilon$ is mainly used for the selection of meta-clusters in classification. $\epsilon$ plays an important role in BPC. Thus, we will make some suggestions on the adjustment of $\epsilon$. The bigger $\epsilon$ value, the smaller the number of patterns that are misclassified, but also the more ambiguous classification results, that is, more patterns belong to meta-clusters. Too small $\epsilon$ value, however, will result in fewer patterns in the meta-cluster, but it may cause more misclassifications for the imprecise patterns. Therefore, $\epsilon$ should be tuned according to the imprecision degree that one can accept under specific situations.

In order to clearly express the basic principle of BPC and facilitate its popularization and application, the main steps are given in Fig. 1.

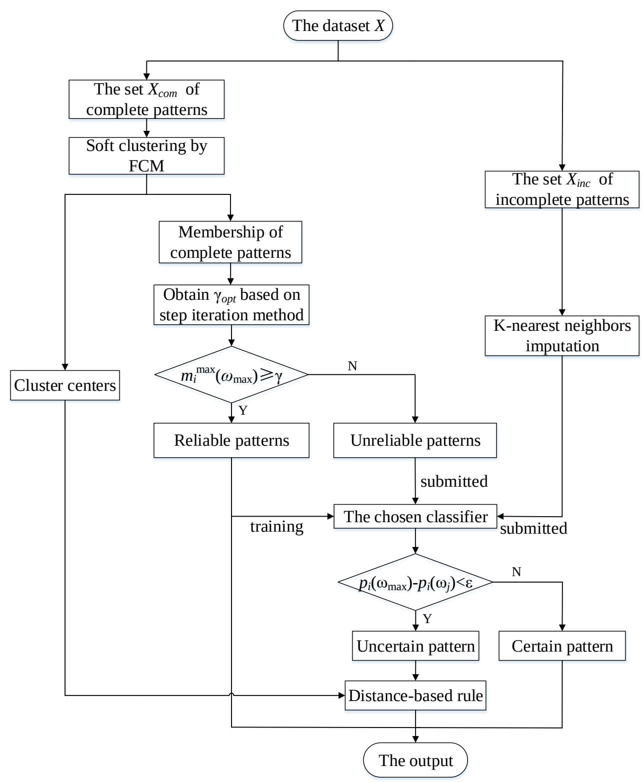



Figure 1: The flowchart of BPC method.

## IV. EXPERIMENT APPLICATIONS

In this section, the performance of BPC will be tested using real datasets and compared with mean imputation (MI) method [15], $K$-nearest neighbors imputation (KNNI) method [17],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



[18], fuzzy $c$-means imputation (FCMI) method [19], [20], optimal whole strategy clustering with probabilistic information granules (POCS) [21], fuzzy-based information decomposition (FID) method [22] and locally linear approximation (LLA) method [23]. In this paper, a relatively simple K-NN classifier is selected as the standard classifier. To fully illustrate that the implementation of BPC does not depend on the chosen classifier, Support Vector Machine (SVM) [41] and Naive Bayesian (NB) [42] methods are also used to test and evaluate the BPC calculation. In addition to the adjusted parameters in the experiments, other parameters are set to default values. Since the introduction of meta-clusters, the error rate ($Re$), imprecision rate ($Ri$), and benefit value are used as indicators [11], [43]. The results in the experiments are the statistical averages of the program execution 10 times.

### A. Experiment 1

Twelve well-known real datasets available from UCI repository [12] to design the experiment. For these datasets, different proportions of incomplete patterns are designed. Each pattern can have a different number (i.e. $\phi$) of missing values, and they are missing completely at random (MCAR). The basic characteristics of the twelve datasets are presented in table I, and all the detailed information can be found at http://archive.ics.uci.edu/ml/.

Table I: Basic Information of The Used Datasets.

| Data | Classes | Attributes | Instances |
|---|---|---|---|
| Iris (Ir) | 3 | 4 | 150 |
| Seeds (Se) | 3 | 7 | 210 |
| Statlog (St) | 2 | 13 | 270 |
| Wine (Wi) | 3 | 13 | 178 |
| Vehicle (Ve) | 4 | 18 | 846 |
| Forest Type (FT) | 4 | 27 | 523 |
| Satimage (Sa) | 7 | 36 | 6435 |
| Segment (Seg) | 7 | 19 | 2310 |
| Texture (Te) | 11 | 40 | 5500 |
| Hayes (Ha) | 3 | 5 | 160 |
| Knowledge (Kn) | 4 | 5 | 403 |
| Abalone (Ab) | 3 | 9 | 4174 |

Fig. 2 (a) - (l) exhibits the optimal threshold for adaptive selection of reliable patterns from the real datasets. For these datasets, we randomly choose 25% of the patterns for missing attributes in diverse dimensions. Here the basic classifier K-NN is adopted to classify the complete patterns where $K = 13$ and $\lambda = 0.01$. One can see that, for most datasets, the optimal threshold $\gamma \in [0.6, 0.95]$ can always be obtained within the membership range (i.e., the error between classification results and FCM clustering results is the smallest). For Forest Type, Satimage, and Segment datasets, the number of classes of reliable patterns will be smaller than that of the complete datasets if the selected threshold is too large, which makes it impossible to train a basic classifier for classification. Through exploratory experiments, however, we can always find an optimal threshold among the previously possible thresholds, thus it will not affect the final results of the proposed method. Also, other basic classifiers (SVM, NB) can be used to classify reliable patterns with adaptive thresholds.

Table II shows the classification results of different methods for the real datasets. We randomly choose $25\%$ of the patterns with missing values in distinct dimensions. One can see clearly from Table II that BPC generally produces lower error rate than the MI, KNNI, FCMI, POCS, FID and LLA methods, but meanwhile it yields some imprecision in the result because of the introduction of meta-cluster, which indicates that some incomplete patterns are hard to be submitted to singleton clusters due to the lack of discriminant information. Although we estimate the missing values, the estimations are not a complete substitute for the truth and inevitably introduce uncertainty and errors. MI, KNNI, FCMI, POCS, FID and LLA all classify patterns into singleton (specific) clusters, which may cause the risk of misclassification for those patterns distributed in the overlapping regions of different singleton clusters or intermediate regions, e.g., due to missing values. By contrast, the BPC submits some uncertain patterns to appropriate meta-clusters in a more cautious way thereby avoiding the misclassification caused by inadequate information as far as possible. With the increases of $\epsilon$ from $\epsilon = 0.2$ to $\epsilon = 0.4$, one can see that the error rate decreases, while the imprecision rate increases. In applications, $\epsilon$ can be adapted according to the imprecision rate one can accept.

Table III illustrates the performance of BPC with different classifiers, and $\epsilon = 0.3$ is the default. One can observe that the chosen classifier in BPC usually produce similar error rates, which is lower than that of other methods, because some "hard-to-classify" patterns are submitted to the relevant meta-clusters automatically and reasonably. This experiment presents that the implementation of BPC does not depend on the selection of the chosen classifiers, and also illustrates the potential of BPC in practical classification tasks. In addition, when the number of missing patterns gradually increases, it usually leads to an increase in the error rate. This is reasonable, because the more attributes that are missing, the greater the uncertainty in pattern classification, and thereby the greater the likelihood of misclassification. In this case, one should more cautiously in classifying the patterns included in meta-clusters.

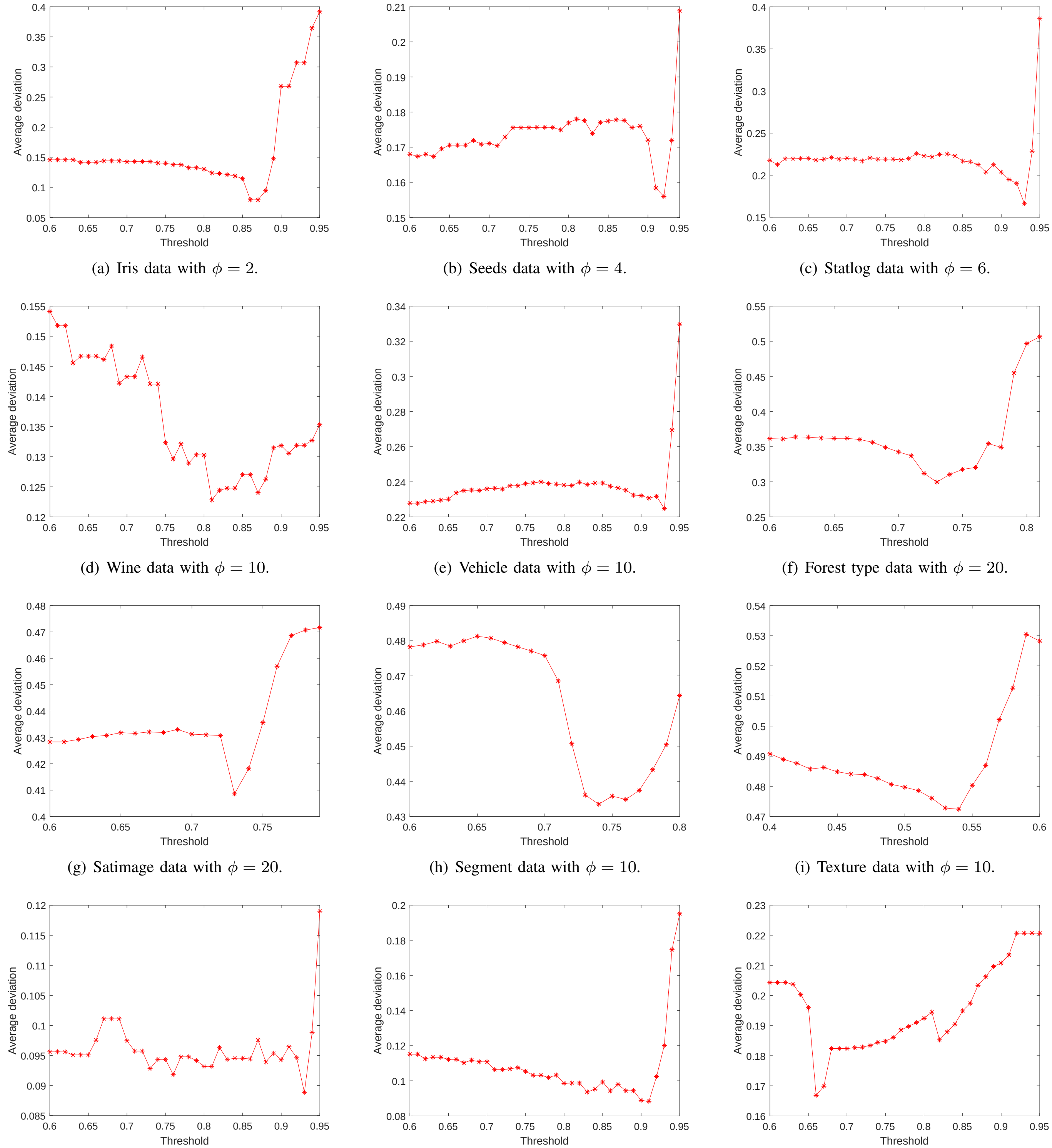
### B. Experiment 2

To further illustrate the effect of different parameters on the performance of the BPC method, we allow $K$ and $\epsilon$ to take different values and the datasets with different missing rates (i.e., 10%, 25%, 40%), respectively.

Fig. 3 (a) - (l) shows more intuitively and clearly the effect of different $K$ values of neighbors on KNNI, LLA, and BPC methods classification results. The abscissa corresponds to $K$ values, ranging from 5 to 15, and the ordinate corresponds to the mean of the error rate in the classification method, denoted by $[0, 1]$. Among them, the nine complete datasets shown in Fig. 3 are missing different dimension attributes in $25\%$ of randomly chose patterns. The error rate of BPC with different classifiers is much lower than that of other methods, and the classification results of different $K$ values change a bit in K-NN, which further illustrates that the classification performance is insensitive to the changing of $K$. This indicates that the BPC method is robust to the choice of $K$ values.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering






(a) Iris data with $\phi = 2$. (b) Seeds data with $\phi = 4$. (c) Statlog data with $\phi = 6$.

(d) Wine data with $\phi = 10$. (e) Vehicle data with $\phi = 10$. (f) Forest type data with $\phi = 20$.

(g) Satimage data with $\phi = 20$. (h) Segment data with $\phi = 10$. (i) Texture data with $\phi = 10$.

(j) Hayes data with $\phi = 2$. (k) Knowledge data with $\phi = 2$. (l) Abalone data with $\phi = 4$.

Figure 2: The optimal threshold $\gamma_{opt}$ obtained by step iteration.

The results in Table IV clearly show that BPC has a lower error rate because it introduces the concept of meta-cluster under belief functions theory. In BPC, if we increase the adjustment parameter from $\epsilon = 0.2$ to $\epsilon = 0.4$, this will lead to a decrease in the error rate but also bring an increase in the imprecision rate. Therefore, in applications, we should choose a suitable compromise between the error rate and the imprecision according to the situation. The results show that the known information is not sufficient to accurately classify the patterns in the meta-clusters. Our tests and analyses illustrate the interest in the unsupervised classification of incomplete data based on belief function theory.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



Table II: The results of the used datasets with different $\epsilon$ values (In %).

| Data | $\phi$ | MI $Re$ | KNNI $Re$ | FCMI $Re$ | POCS $Re$ | FID $Re$ | LLA $Re$ | BPC($\epsilon$=0.2) $\{Re, Ri\}$ | BPC($\epsilon$=0.3) $\{Re, Ri\}$ | BPC($\epsilon$=0.4) $\{Re, Ri\}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Ir | 1 | 14.00 | 10.00 | 10.00 | 10.67 | 12.67 | 10.00 | {**8.67**, 3.33} | {**6.00**, 6.00} | {**4.67**, 8.00} |
| | 2 | 15.33 | 10.67 | 10.00 | 11.33 | 16.00 | 10.00 | {**9.33**, 2.67} | {**8.67**, 3.33} | {**6.67**, 6.67} |
| | 3 | 16.67 | 11.33 | 12.00 | 12.00 | 16.00 | 10.67 | {**10.00**, 2.67} | {**8.67**, 5.33} | {**7.33**, 6.67} |
| Se | 2 | 13.33 | 10.48 | 10.48 | 10.48 | 14.29 | 10.95 | {**9.05** , 0.95} | {**7.14**, 5.71} | {**6.67**, 7.62} |
| | 4 | 20.00 | 11.43 | 12.38 | 13.33 | 19.52 | 11.90 | {**10.95**, 2.38} | {**8.10**, 5.71} | {**8.10**, 8.10} |
| | 6 | 24.76 | 16.19 | 15.71 | 17.38 | 27.14 | 20.48 | {**13.81**, 5.24} | {**12.38**, 7.62} | {**9.52**, 12.86} |
| St | 3 | 41.48 | 40.00 | 40.74 | 39.26 | 40.74 | 41.85 | {**38.52**, 2.59} | {**37.41**, 6.30} | {**34.44**, 10.00} |
| | 6 | 42.59 | 42.59 | 42.59 | 41.58 | 42.22 | 42.59 | {**40.37**, 3.70} | {**37.78**, 7.78} | {**37.04**, 10.00} |
| | 10 | 42.96 | 43.33 | 42.96 | 42.96 | 41.11 | 44.07 | {**41.11**, 5.19} | {**38.52**, 10.00} | {**37.41**, 12.22} |
| Wi | 3 | 33.71 | 32.02 | 34.27 | 34.27 | 34.27 | 33.15 | {**30.90**, 1.12} | {**29.78**, 3.93} | {**27.53**, 6.74} |
| | 6 | 37.64 | 34.27 | 35.39 | 35.96 | 38.20 | 35.39 | {**32.02**, 0.56} | {**29.78**, 3.93} | {**28.65**, 6.74} |
| | 10 | 44.94 | 37.08 | 38.76 | 39.89 | 44.38 | 41.57 | {**36.52**, 2.25} | {**33.71**, 6.74} | {**33.15**, 10.11} |
| Ve | 5 | 57.68 | 55.08 | 55.91 | 61.58 | 54.96 | 55.08 | {**53.43**, 2.72} | {**52.60**, 3.90} | {**52.01**, 4.96} |
| | 10 | 57.68 | 55.44 | 57.45 | 62.36 | 57.21 | 55.44 | {**54.61**, 2.01} | {**53.31**, 4.02} | {**52.25**, 5.44} |
| | 15 | 59.69 | 56.38 | 58.39 | 62.70 | 65.96 | 56.74 | {**54.61**, 3.66} | {**53.55**, 5.20} | {**52.36**, 6.62} |
| FT | 5 | 21.80 | 22.18 | 21.80 | 22.37 | 23.71 | 21.99 | {**20.46**, 3.25} | {**19.12**, 6.12} | {**17.40**, 8.41} |
| | 10 | 22.18 | 22.37 | 21.99 | 22.60 | 25.62 | 22.18 | {**21.41**, 2.68} | {**19.31**, 5.16} | {**18.36**, 7.46} |
| | 20 | 34.80 | 22.56 | 33.84 | 24.13 | 33.08 | 22.56 | {**22.37**, 2.87} | {**20.65**, 4.78} | {**19.69**, 6.69} |
| Sa | 10 | 32.73 | 30.63 | 32.39 | 30.86 | 33.13 | 30.63 | {**28.17**, 1.60} | {**27.40**, 2.81} | {**26.81**, 3.85} |
| | 20 | 40.00 | 30.80 | 39.18 | 30.94 | 39.08 | 30.89 | {**28.36**, 1.60} | {**27.52**, 2.80} | {**26.84**, 3.89} |
| | 30 | 52.40 | 31.02 | 51.64 | 33.46 | 52.88 | 31.33 | {**28.55**, 1.68} | {**27.82**, 2.74} | {**27.20**, 3.89} |
| Seg | 5 | 42.34 | 37.84 | 45.63 | 39.22 | 52.73 | 40.35 | {**35.28**, 3.16} | {**33.81**, 5.37} | {**32.64**, 7.32} |
| | 10 | 51.39 | 38.53 | 51.39 | 42.42 | 54.81 | 38.74 | {**36.93**, 2.73} | {**35.76**, 4.85} | {**33.72**, 7.45} |
| | 15 | 54.29 | 42.08 | 58.14 | 47.36 | 54.63 | 42.94 | {**38.57**, 2.81} | {**37.62**, 4.68} | {**35.11**, 7.84} |
| Te | 5 | 39.02 | 38.09 | 39.07 | 35.80 | 39.05 | 38.09 | {**31.70**, 2.55} | {**31.15**, 3.58} | {**30.35**, 4.78} |
| | 10 | 41.35 | 38.15 | 40.87 | 38.35 | 47.22 | 38.09 | {**34.16**, 2.09} | {**33.49**, 3.13} | {**32.45**, 4.62} |
| | 30 | 53.65 | 38.47 | 53.11 | 51.65 | 56.60 | 38.65 | {**34.22**, 2.31} | {**33.64**, 3.27} | {**32.51**, 5.07} |
| Ha | 1 | 56.06 | 56.82 | 56.06 | 56.06 | 53.79 | 58.33 | {**53.79**, 4.55} | {**52.27**, 8.33} | {**50.76**, 12.12} |
| | 2 | 61.36 | 60.61 | 61.36 | 60.61 | 59.75 | 60.61 | {**57.68**, 4.55} | {**55.30**, 7.58} | {**54.55**, 10.61} |
| | 3 | 61.36 | 60.61 | 61.36 | 62.12 | 62.12 | 61.36 | {**59.09**, 2.27} | {**58.33**, 6.82} | {**56.06**, 11.36} |
| Kn | 1 | 56.82 | 55.30 | 56.82 | 57.58 | 56.82 | 58.33 | {**53.79**, 4.55} | {**51.52**, 9.09} | {**50.00**, 12.88} |
| | 2 | 59.85 | 59.85 | 59.85 | 59.09 | 57.58 | 60.61 | {**54.55**, 5.30} | {**52.27**, 10.61} | {**52.27**, 12.88} |
| | 3 | 60.61 | 59.85 | 60.61 | 59.85 | 61.36 | 61.36 | {**56.06**, 4.55} | {**54.55**, 7.58} | {**51.52**, 12.88} |
| Ab | 2 | 49.35 | 48.59 | 48.25 | 48.90 | 49.59 | 48.54 | {**46.14**, 3.35} | {**44.11**, 6.71} | {**42.02**, 10.04} |
| | 4 | 50.50 | 48.80 | 50.77 | 49.64 | 50.69 | 49.31 | {**46.21**, 3.64} | {**44.39**, 6.61} | {**42.48**, 9.87} |
| | 6 | 51.77 | 48.66 | 50.91 | 52.16 | 51.65 | 49.50 | {**46.86**, 2.06} | {**45.50**, 4.46} | {**44.11**, 6.73} |

In order to show the performance of BPC more intuitively, we employ the benefit value [43] to evaluate the results of all methods, which is an index that can balance the error and imprecision. It is interesting to note that when no patterns are submitted to meta-clusters, the benefit value is equal to the accuracy. the imprecision penalizing coefficient is recommended to be within [0,1], we take the penalizing coefficient of 0.5 to test the effectiveness of BPC with respect to other related methods. K-NN is chosen as the basic classifier. One can see from Fig. 4 that the BPC generally yields higher benefit value than other methods, which indicates that the imprecision in meta-clusters is better than error in decision-making. The BPC results clearly show that the attributes used in the real datasets are not enough to correctly classify the patterns in meta-clusters. We should be more cautious about these patterns and need additional sources of information to achieve better concrete results (if necessary).

## V. CONCLUSION

In this paper, we propose a new belief-based incomplete pattern unsupervised classification method (BPC), which aims to reduce the negative effect of estimations on clustering and characterize the imprecision in the results caused by missing values. In BPC, we first obtain the real cluster centers by soft clustering and thereby obtain the reliable patterns by a step iteration method. Then the unreliable patterns and the edited incomplete patterns are classified by a basic classifier trained by the reliable patterns and thereby are regarded as certain or uncertain ones. For uncertain patterns, a new distance-based rule is introduced to further extract the cluster information, and a few patterns may be submitted to meta-clusters which can characterize the uncertainty and imprecision caused by missing values and, from a cautious perspective, thereby reduce the risk of misclassification. If one wants more accurate results, some other (possibly expensive) technology or information source should be introduced. The performance of BPC and other classical methods is evaluated by using eleven datasets of UCI repository. The results show that BPC can effectively reduce the error rates and reasonably characterize the uncertainty and imprecision caused by the missing attribute information.

In our work, we use single imputation based on the KNNI method to estimate incomplete patterns, however, there is a problem that the neighbors may belong to multiple clusters. In the future, we will consider the correlation between attributes

Table III: The results of the used datasets with different classifiers (In %).

| Data | $\phi$ | MI | KNNI | FCMI | POCS | FID | LLA | $\text{BPC}_{\text{KNN}}$ | $\text{BPC}_{\text{SVM}}$ | $\text{BPC}_{\text{NB}}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $Re$ | $Re$ | $Re$ | $Re$ | $Re$ | $Re$ | $\{Re, Ri\}$ | $\{Re, Ri\}$ | $\{Re, Ri\}$ |
| Ir | 1 | 14.00 | 10.00 | 10.00 | 10.67 | 12.67 | 10.00 | {**6.00**, 6.00} | {**4.67**, 9.33} | {**6.67**, 4.00} |
| | 2 | 15.33 | 10.67 | 10.00 | 11.33 | 16.00 | 10.00 | {**8.67**, 3.33} | {**7.33**, 7.33} | {**8.67**, 0.67} |
| | 3 | 16.67 | 11.33 | 12.00 | 12.00 | 16.00 | 10.67 | {**8.67**, 5.33} | {**7.33**, 8.00} | {**10.00**, 0.67} |
| Se | 2 | 13.33 | 10.48 | 10.48 | 10.48 | 14.29 | 10.95 | {**7.14**, 5.71} | {**5.71**, 10.00} | {**8.57**, 1.43} |
| | 4 | 20.00 | **11.43** | 12.38 | 13.33 | 19.52 | 11.90 | {**8.10**, 5.71} | {**6.67**, 11.43} | {**11.43**, 0.48} |
| | 6 | 24.76 | 16.19 | 15.71 | 17.38 | 27.14 | 20.48 | {**12.38**, 7.62} | {**8.57**, 14.76} | {**13.81**, 0.48} |
| St | 3 | 41.48 | 40.00 | 40.74 | 39.26 | 40.74 | 41.85 | {**37.41**, 6.30} | {**32.22**, 13.70} | {**35.56**, 4.07} |
| | 6 | 42.59 | 42.59 | 42.59 | 41.58 | 42.22 | 42.59 | {**37.78**, 7.78} | {**34.81**, 14.07} | {**37.78**, 3.70} |
| | 10 | 42.96 | 43.33 | 42.96 | 42.96 | 41.11 | 44.07 | {**38.52**, 10.00} | {**34.44**, 16.30} | {**38.89**, 5.93} |
| Wi | 3 | 33.71 | 32.02 | 34.27 | 34.27 | 34.27 | 33.15 | {**29.78**, 3.93} | {**26.97**, 6.18} | {**27.53**, 1.12} |
| | 6 | 37.64 | 34.27 | 35.39 | 35.96 | 38.20 | 35.39 | {**29.78**, 3.93} | {**28.09**, 7.30} | {**28.09**, 1.12} |
| | 10 | 44.94 | 37.08 | 38.76 | 39.89 | 44.38 | 41.57 | {**33.71**, 6.74} | {**30.34**, 11.24} | {**32.02**, 1.12} |
| Ve | 5 | 57.68 | 55.08 | 55.91 | 61.58 | 54.96 | 55.08 | {**52.60**, 3.90} | {**48.11**, 10.87} | {**54.37**, 0.71} |
| | 10 | 57.68 | 55.44 | 57.45 | 62.36 | 57.21 | 55.44 | {**53.31**, 4.02} | {**49.05**, 10.52} | {**55.67**, 0.47} |
| | 15 | 59.69 | 56.38 | 58.39 | 62.70 | 65.96 | 56.74 | {**53.55**, 5.20} | {**49.41**, 11.47} | {**55.67**, 0.71} |
| FT | 5 | 21.80 | 22.18 | 21.80 | 22.37 | 23.71 | 21.99 | {**19.12**, 6.12} | {**14.53**, 13.77} | {**17.59**, 0} |
| | 10 | 22.18 | 22.37 | 21.99 | 22.60 | 25.62 | 22.18 | {**19.31**, 5.16} | {**14.72**, 12.81} | {**17.59**, 0.57} |
| | 20 | 34.80 | 22.56 | 33.84 | 24.13 | 33.08 | 22.56 | {**20.65**, 4.78} | {**15.30**, 14.53} | {**17.59**, 0.96} |
| Sa | 10 | 32.73 | 30.63 | 32.39 | 30.86 | 33.13 | 30.63 | {**27.40**, 2.81} | {**22.87**, 14.65} | {**28.72**, 0.12} |
| | 20 | 40.00 | 30.80 | 39.18 | 30.94 | 39.08 | 30.89 | {**27.52**, 2.80} | {**22.91**, 15.09} | {**28.78**, 0.20} |
| | 30 | 52.40 | 31.02 | 51.64 | 33.46 | 52.88 | 31.33 | {**27.82**, 2.74} | {**22.94**, 14.51} | {**29.01**, 0.06} |
| Seg | 5 | 42.34 | 37.84 | 45.63 | 39.22 | 52.73 | 40.35 | {**33.81**, 5.37} | {**28.05**, 15.63} | \ |
| | 10 | 51.39 | 38.53 | 51.39 | 42.42 | 54.81 | 38.74 | {**35.76**, 4.85} | {**29.09**, 15.50} | \ |
| | 15 | 54.29 | 42.08 | 58.14 | 47.36 | 54.63 | 42.94 | {**37.62**, 4.68 } | {**29.61**, 16.88} | \ |
| Te | 5 | 39.02 | 38.09 | 39.07 | **35.80** | 39.05 | 38.09 | {**31.15**, 3.58} | {**29.11**, 12.96} | {36.41, 0.05} |
| | 10 | 41.35 | 38.15 | 40.87 | 38.35 | 47.22 | 38.09 | {**33.49**, 3.13} | {**29.15**, 13.60} | {**37.16**, 0.22} |
| | 30 | 53.65 | 38.47 | 53.11 | 51.65 | 56.60 | 38.65 | {**33.64**, 3.27} | {**29.42**, 13.38} | {**37.42**, 0.20} |
| Ha | 1 | 56.06 | 56.82 | 56.06 | 56.06 | 53.79 | 58.33 | {**52.27**, 8.33} | {**53.03**, 6.82} | {**53.03**, 3.03} |
| | 2 | 61.36 | 60.61 | 61.36 | 60.61 | 59.75 | 60.61 | {**55.30**, 7.58} | {**53.79**, 11.36} | {**56.06**, 4.55} |
| | 3 | 61.36 | 60.61 | 61.36 | 62.12 | 62.12 | 61.36 | {**58.33**, 6.82} | {**56.82**, 9.09} | {**59.09**, 1.52} |
| Kn | 1 | 56.82 | 55.30 | 56.82 | 57.58 | 56.82 | 58.33 | {**51.52**, 9.09} | {**49.24**, 11.36} | {**53.79**, 4.55} |
| | 2 | 59.85 | 59.85 | 59.85 | 59.09 | 57.58 | 60.61 | {**52.27**, 10.61} | {**53.03**, 9.85} | {**52.27**, 3.03} |
| | 3 | 60.61 | 59.85 | 60.61 | 59.85 | 61.36 | 61.36 | {**54.55**, 7.58} | {**53.79**, 10.61} | {**56.82**, 5.3} |
| Ab | 2 | 49.35 | 48.59 | 48.25 | 48.90 | 49.59 | 48.54 | {**44.11**, 6.71} | {**38.67**, 16.12} | {**45.60**, 2.40} |
| | 4 | 50.50 | 48.80 | 50.77 | 49.64 | 50.69 | 49.31 | {**44.39**, 6.61} | {**40.99**, 12.87} | {**46.14**, 2.56} |
| | 6 | 51.77 | 48.66 | 50.91 | 52.16 | 51.65 | 49.50 | {**45.50**, 4.46} | {**39.17**, 15.57} | {**46.43**, 1.73} |

and adopt multiple imputation strategies to reduce the negative effect of misestimations as much as possible. In the long-term, we hope to establish an adaptive estimation strategy in the clustering task and to reasonably characterize the uncertainty and imprecision caused by missing values.

## References


[1] J. Liang, L. Bai, C. Dang. "The $k$-means-type algorithms versus imbalanced data distributions", IEEE Trans. Fuzzy Syst., vol. 20, no. 4, pp. 728−745, 2012.

[2] Y. Zhang, S. Chen and G. Yu. "Efficient distributed density peaks for clustering large data sets in MapReduce", IEEE Trans. Knowledge and Data Eng., vol. 28, no. 12, pp. 3218−3230, 2016.

[3] V. D'Orangeville, M. A. Mayers. "Efficient cluster labeling for support vector clustering", IEEE Trans. Knowledge and Data Eng., vol. 25, no. 11, pp. 2494−2506, 2013.

[4] H. Liu, J. Wu, T. Liu. "Spectral ensemble clustering via weighted $k$-means: theoretical and practical evidence", IEEE Trans. Knowledge and Data Eng., vol. 29, no. 5, pp. 1129−1143, 2017.

[5] A. K. Jain. "Data clustering: 50 years beyond $k$-means", Pattern recognit. lett., vol. 31, no. 8, pp. 651−666, 2010.

[6] J. C. Bezdek. "Pattern recognition with fuzzy objective function algorithms", Adv. Appl. Pattern Recognit., vol. 22, no. 1171, pp. 203−239, 1981.

[7] J. Nayak, B. Naik, H. S. Behera. "Fuzzy $c$-means (FCM) clustering algorithm: a decade review from 2000 to 2014", Comput. Intell. Data Mining, vol. 2, 2015.

[8] R. Krishnapuram, J. Keller. "The possibilistic $c$-means algorithm: insights and recommendations", IEEE Trans. Fuzzy Syst., vol. 4, no. 3, pp. 385−393, 1996.

[9] R. Dave. "Clustering of relational data containing noise and outliers", Pattern Recognit. Lett., no. 12, pp. 657−664, 1991.

[10] M. H. Masson and T. Denœux. "ECM: An evidential version of the fuzzy $c$-means algorithm", Pattern Recognit., vol. 41, no. 4, pp. 1384−1397, 2008.

[11] Z. -G. Liu, Q. Pan, J. Dezert. "Credal $c$-means clustering method based on belief functions", Knowl.-Based Syst., vol. 74, no. 1, pp. 119−132, 2015.

[12] M. Lichman. UCI machine learning repository, (2013). URL http://archive.ics.uci.edu/ml.

[13] P. García-Laencina, J. Sancho-Gómez, A. Figueiras-Vidal. "Pattern classification with missing data: A review", Neural Comput. Appl. vol. 19, no. 2, pp. 263−282, 2010.

[14] R. J. Little, D. B. Rubin. "Statistical Analysis With Missing Data", 2nd ed. New York, NY, USA: Wiley, 2002.

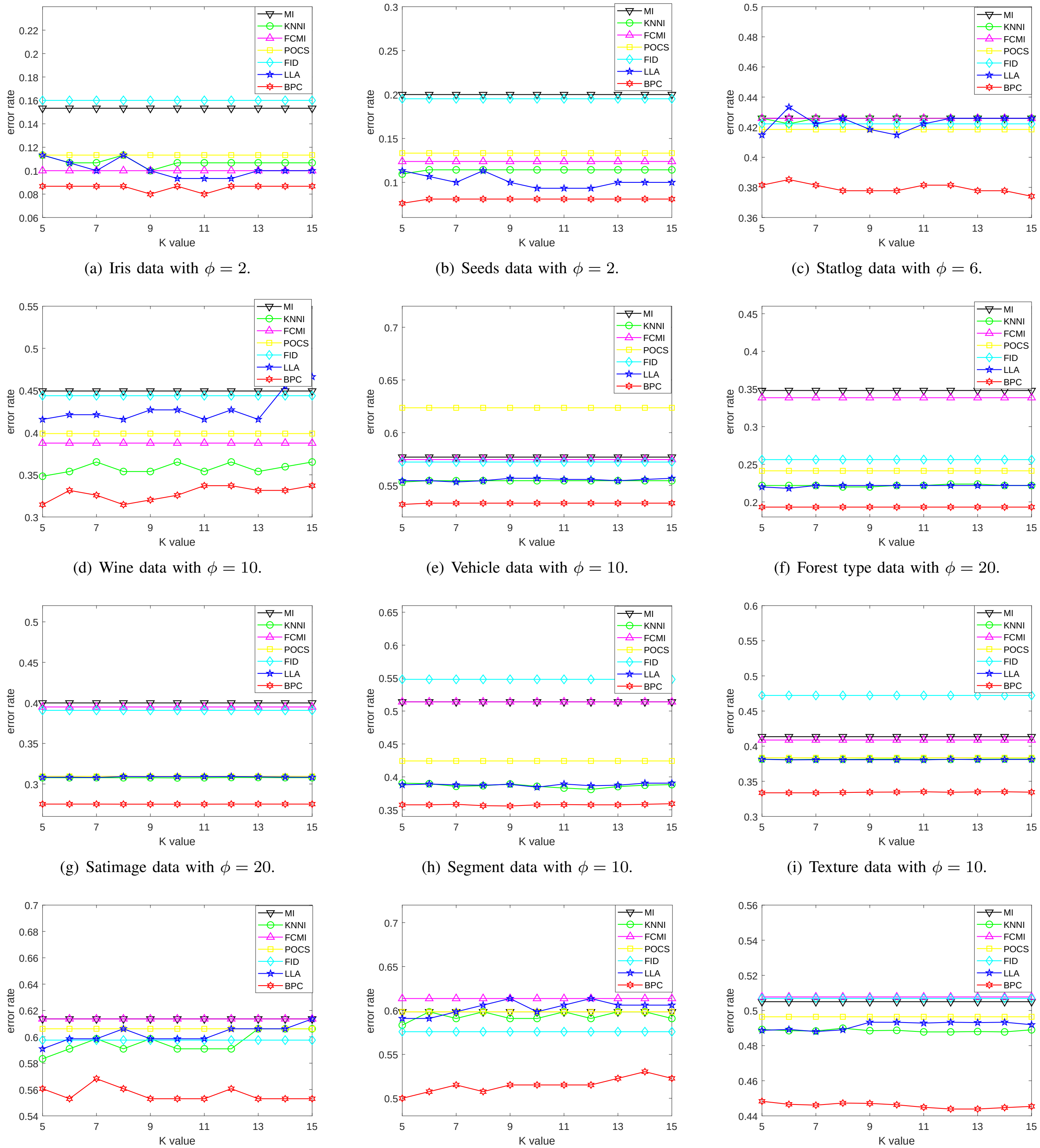
[15] J. L. Schafer. "Analysis of incomplete multivariate data", Chapman & Hall, Flor., 1997.

[16] G. Batista, M. C. Monard. "A study of $k$-nearest neighbour as an imputation method", Proc. 2nd Int. Conf. Hybrid Intell. Syst., vol. 87, pp. 251−260, 2001.

[17] M. Zhu, X. -B. Cheng. "Iterative KNN imputation based on GRA for missing values in TPLMS", Proc. 4th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT), pp. 94−99, 2015.

[18] L. P. Brás, J. C. Menezes. "Improving cluster-based missing value estimation of DNA microarray data", Biomol. Engineering, vol. 24, no. 2, pp. 273−282, 2007.

[19] J. Luengo, J. A. Saez, and F. Herrera. "Missing data imputation for




(a) Iris data with $\phi = 2$.
(b) Seeds data with $\phi = 2$.
(c) Statlog data with $\phi = 6$.
(d) Wine data with $\phi = 10$.
(e) Vehicle data with $\phi = 10$.
(f) Forest type data with $\phi = 20$.
(g) Satimage data with $\phi = 20$.
(h) Segment data with $\phi = 10$.
(i) Texture data with $\phi = 10$.
(j) Hayes data with $\phi = 2$.
(k) Knowledge data with $\phi = 2$.
(l) Abalone data with $\phi = 4$.

Figure 3: The results of the used datasets with different $K$ values.


fuzzy rule-based classification systems", Soft Comput., vol. 16, no. 5, pp. 863−881, 2012.

[20] D. Li, J. Deogun, W. Spaulding, and B. Shuart. "Towards missing data imputation: A study of fuzzy $k$-means clustering method", Proc. 4th Int. Conf. Rough Sets Current Trends Comput. (RSCTC04), Uppsala, Sweden, pp. 573−579, 2004.

[21] L. Zhang, W. Lu, X. Liu. "Fuzzy $c$-means clustering of incomplete data based on probabilistic information granules of missing values", Knowl.-Based Syst., vol. 99, pp. 51−70, 2016.

[22] Liu, S., Zhang, J., Xiang, Y., Zhou, W.. "Fuzzy-based information decomposition for incomplete and imbalanced data learning", IEEE Trans. Fuzzy Syst., vol. 25, no. 6, pp. 1476-1490, 2017.

[23] J. H. Dai, H. Hu, Q. H. Hu. "Locally linear approximation approach for incomplete data", IEEE Trans. Cybern., vol. 48, no. 6, pp. 1720-1732, 2018.

[24] T. Denoeux. "Maximum likelihood estimation from uncertain data in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



Table IV: Thre results of the used datasets with different missing rates under different $\epsilon$ (In %).

| Data | $\phi$ | missing % | MI $Re$ | KNNI $Re$ | FCMI $Re$ | POCS $Re$ | FID $Re$ | LLA $Re$ | BPC($\epsilon$=0.2) $\{Re, Ri\}$ | BPC($\epsilon$=0.3) $\{Re, Ri\}$ | BPC($\epsilon$=0.4) $\{Re, Ri\}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ir | 2 | 10 | 11.33 | 8.67 | 8.67 | 8.67 | 12.00 | 8.00 | {**7.33**, 2.67} | {**5.33**, 6.67} | {**4.67**, 7.33} |
| | | 25 | 15.33 | 10.67 | 10.00 | 11.33 | 16.00 | 10.00 | {**9.33**, 2.67} | {**8.67**, 3.33} | {**6.67**, 6.67} |
| | | 40 | 18.67 | 13.33 | 13.33 | 12.92 | 18.67 | 14.00 | {**10.00**,5.33} | {**8.67**,8.00} | {**7.33**,10.67} |
| Se | 4 | 10 | 14.29 | 9.52 | 10.95 | 10.76 | 15.71 | 10.00 | {**8.10**, 4.29} | {**7.62**, 5.24} | {**7.14**, 7.14} |
| | | 25 | 20.00 | 11.43 | 12.38 | 13.33 | 19.52 | 11.90 | {**10.95**, 2.38} | {**8.10**, 5.71} | {**8.10**, 8.10} |
| | | 40 | 27.62 | 15.24 | 17.62 | 17.86 | 21.90 | 15.24 | {**12.38**, 4.29} | {**11.90**, 6.19} | {**10.00**, 9.52} |
| St | 6 | 10 | 40.37 | 40.37 | 40.74 | 40.37 | 41.11 | 40.74 | {**37.78**, 3.33} | {**37.04**, 5.93} | {**36.30**, 7.41} |
| | | 25 | 42.59 | 42.59 | 42.59 | 41.58 | 42.22 | 42.59 | {**40.37**, 3.70} | {**37.78**, 7.78} | {**37.04**, 10.00} |
| | | 40 | 44.44 | 41.85 | 44.07 | 42.22 | 44.81 | 43.33 | {**40.74**, 4.81} | {**38.52**, 8.89} | {**37.04**, 12.22} |
| Wi | 6 | 10 | 34.27 | 30.34 | 32.58 | 32.02 | 33.15 | 30.34 | {**29.21**, 2.25} | {**29.21**, 2.81} | {**26.40**, 7.30} |
| | | 25 | 44.94 | 37.08 | 38.76 | 39.89 | 44.38 | 41.57 | {**36.52**, 2.25} | {**33.71**, 6.74} | {**33.15**, 10.11} |
| | | 40 | 47.75 | 40.45 | 42.13 | 41.76 | 42.70 | 40.45 | {**38.20**, 5.06} | {**35.39**, 11.80} | {**34.83**, 12.92} |
| Ve | 10 | 10 | 56.86 | 54.61 | 55.79 | 61.58 | 55.56 | 54.73 | {**53.43**, 2.72} | {**52.48**, 3.78} | {**51.77**, 4.61} |
| | | 25 | 57.68 | 55.44 | 57.45 | 62.36 | 57.21 | 55.44 | {**54.61**, 2.01} | {**53.31**, 4.02} | {**52.25**, 5.44} |
| | | 40 | 66.08 | 56.15 | 58.51 | 62.41 | 60.76 | 55.56 | {**55.20**, 1.06} | {**54.26**, 2.13} | {**53.43**, 3.55} |
| FT | 10 | 10 | 24.67 | 21.03 | 24.47 | 22.33 | 24.09 | **20.84** | {21.99, 1.72} | {21.03, 3.44} | {**18.74**, 7.65} |
| | | 25 | 34.80 | 22.56 | 33.84 | 24.13 | 33.08 | 22.56 | {**21.41**, 3.25} | {**19.69**, 5.35} | {**18.93**, 6.88} |
| | | 40 | 49.71 | 24.09 | 47.23 | 26.54 | 37.48 | 23.71 | {**22.37**, 3.82} | {**21.61**, 5.54} | {**19.89**, 8.03} |
| Sa | 20 | 10 | 34.76 | 30.64 | 34.25 | 30.72 | 34.09 | 30.72 | {**28.10**, 1.35} | {**27.26**, 2.53} | {**26.45**, 3.85} |
| | | 25 | 40.00 | 30.80 | 39.18 | 30.94 | 39.08 | 30.89 | {**28.36**, 1.60} | {**27.52**, 2.80} | {**26.84**, 3.89} |
| | | 40 | 51.93 | 30.88 | 51.30 | 31.36 | 51.90 | 30.97 | {**28.24**, 1.62} | {**27.20**, 3.01} | {**26.39**, 4.21} |
| Seg | 10 | 10 | 40.69 | 38.01 | 40.78 | 40.04 | 51.90 | 37.10 | {**36.32**, 2.55} | {**34.94**, 4.46} | {**32.99**, 7.32} |
| | | 25 | 51.39 | 38.53 | 51.39 | 42.42 | 54.81 | 38.74 | {**36.93**, 2.73} | {**35.76**, 4.85} | {**33.72**, 7.45} |
| | | 40 | 54.42 | 41.65 | 55.02 | 48.54 | 55.24 | 41.56 | {**37.92**, 4.63} | {**35.80**, 7.44} | {**34.99**, 8.61} |
| Te | 10 | 10 | 39.20 | 38.00 | 38.98 | 36.00 | 37.35 | 38.05 | {**32.20**, 2.80} | {**31.47**, 3.95} | {**30.67**, 5.16} |
| | | 25 | 41.35 | 38.15 | 40.87 | 38.35 | 47.22 | 38.09 | {**34.16**, 2.09} | {**33.49**, 3.13} | {**32.45**, 4.62} |
| | | 40 | 46.47 | 37.91 | 45.16 | 45.20 | 53.44 | 37.87 | {**35.04**, 1.82} | {**34.13**, 3.31} | {**32.75**, 5.71} |
| Ha | 2 | 10 | 58.33 | 57.58 | 58.33 | 57.58 | **56.06** | 57.58 | {56.82, 3.03} | {**53.79**, 7.58} | {**50.76**, 12.12} |
| | | 25 | 61.36 | 60.61 | 61.36 | 60.61 | 59.75 | 60.61 | {**57.68**, 4.55} | {**55.30**, 7.58} | {**54.55**, 10.61} |
| | | 40 | 63.64 | 61.36 | 62.12 | 61.36 | 59.85 | 63.64 | {**58.33**, 6.06} | {**55.30**, 12.12} | {**54.55**, 12.88} |
| Kn | 2 | 10 | 56.82 | 57.58 | 56.82 | 58.33 | 57.58 | 56.82 | {**53.79**, 3.79} | {**52.27**, 6.82} | {**49.24**, 11.36} |
| | | 25 | 59.85 | 59.85 | 59.85 | 59.09 | 57.58 | 60.61 | {**54.55**, 5.30} | {**52.27**, 10.61} | {**52.27**, 12.88} |
| | | 40 | 59.09 | 62.88 | 61.36 | 59.85 | 57.58 | 61.36 | {**56.82**, 5.30} | {**54.55**, 10.61} | {**53.03**, 12.88} |
| Ab | 4 | 10 | 48.40 | 48.63 | 49.19 | 49.11 | 49.38 | 48.75 | {**46.17**, 3.59} | {**44.20**, 6.88} | {**42.43**, 9.99} |
| | | 25 | 50.50 | 48.80 | 50.77 | 49.64 | 50.69 | 49.31 | {**46.21**, 3.64} | {**44.39**, 6.61} | {**42.48**, 9.87} |
| | | 40 | 52.16 | 48.95 | 51.03 | 50.34 | 51.96 | 49.40 | {**47.20**, 2.00} | {**45.78**, 4.31} | {**44.25**, 6.88} |


the belief function framework", IEEE Trans. on Knowledge and Data Eng., vol. 25, no. 1, pp. 119−130, 2013.

[25] X. Zhu, S. Zhang, Z. Jin, Z. Zhang and Z. Xu. "Missing value estimation for mixed-attribute data sets", IEEE Trans. Knowledge and Data Eng., vol. 23, no. 1, pp. 110−121, 2011.

[26] R. J. Hathaway, J. C. Bezdek. "Fuzzy *c*-means clustering of incomplete data", IEEE Trans. Syst. Man Cybern. B Cybern., vol. 31, no. 5, pp. 735−744, 2001.

[27] Z. Zhang, H. Fang. "Multiple− vs non− or single−imputation based fuzzy clustering for incomplete longitudinal behavioral intervention data", IEEE 1st Int. Conf. Connect. Heal.: Appl., 2016.

[28] D. B. Rubin. "Multiple imputation for nonresponse in surveys", NewYork, NY, USA: Wiley, 1987.

[29] G. Shafer. "A mathematical theory of evidence". Princeton, NJ, USA: Princeton Univ. Press, 1976.

[30] P. Smets. "The combination of evidence in the transferable belief model", IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 5, pp. 447−458, 1990.

[31] T. Denœux and M. H. Masson. "EVCLUS: evidential clustering of proximity data", IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 1, pp. 95−109, 2004.

[32] T. Denœux, S. Sriboonchitta, O. Kanjanatarakul. "Evidential clustering of large dissimilarity data", Knowl.-Based Syst., vol. 106, pp. 179−195, 2016.

[33] Z. -G. Su and T. Denœux. "BPEC: Belief-peaks evidential clustering", IEEE Trans. Fuzzy Syst., vol. 27, no. 1, pp. 111−123, 2019.

[34] A. Rodriguez and A. Laio. "Clustering by fast search and find of density peaks", Science, vol. 344, pp. 1492−1496, 2014.

[35] C. Lian, S. Ruan and T. Denœux. "Dissimilarity metric learning in the belief function framework", IEEE Trans. Fuzzy Syst., vol. 24, no. 6, pp. 1555−1564, 2016.

[36] Z. F. Ma, H. P. Tian, Z. C. Liu, Z. W. Zhang. "A new incomplete pattern belief classification method with multiple estimations based on KNN", Applied Soft Comput., vol. 90, 2020. doi.org/10.1016/j.asoc.2020.106175.

[37] X. Xu, J. Zheng, J. -B. Yang. "Data classification using evidence reasoning rule", Knowl.-Based Syst., vol. 116, 2016.

[38] D. Han, J. Dezert and Z. Duan. "Evaluation of probability transformations of belief functions for decision making", IEEE Trans. Syst. Man, Cyber. Syst., vol. 46, no. 1, pp. 93−108, 2016.

[39] T. Denœux. "Decision-making with belief functions: a review", Int. J. Approx. Reason., vol. 109, pp. 87−110, 2019.

[40] S. Zhang, X. Li, M. Zong. "Efficient kNN classification with different numbers of nearest neighbors", IEEE Trans. on Neural Netw. Learn. Syst., vol. 29, no. 5, pp. 1774−1785, 2018.

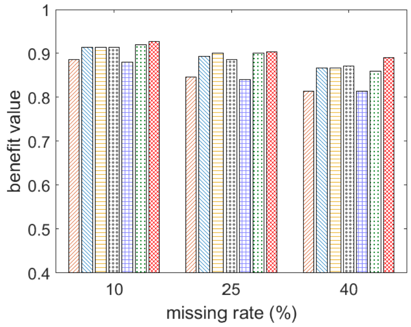
[41] A. Mathur and G. M. Foody. "Multiclass and binary SVM classification: Iimplications for training and classification users", IEEE Geosci. Remote Sens., vol. 5, no. 2, pp. 241−245, 2008.

[42] S. Russell and P. Norvig. "Artificial intelligence: a modern approach", Prentice Hall, 2003.

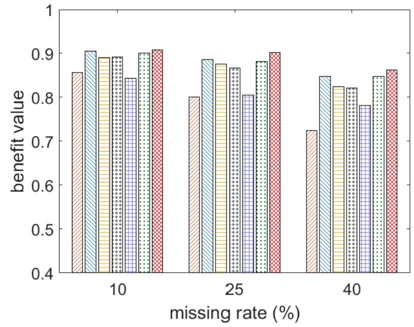
[43] Z. G. Liu, Q. Pan, J. Dezert, A. Martin. "Combination of classifiers with optimal weight based on evidential reasoning", IEEE Trans. Fuzzy Syst., vol. 26, no. 3, pp. 1217−1230, 2018.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering

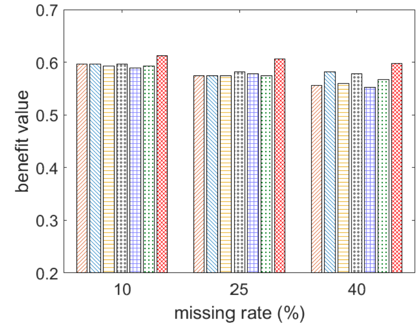




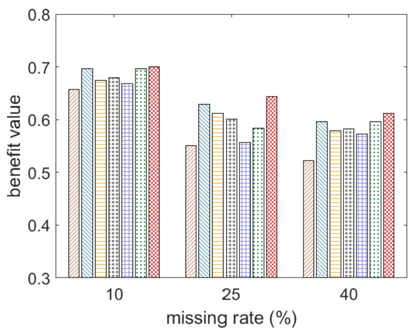

(a) Iris data with $\phi = 2$.



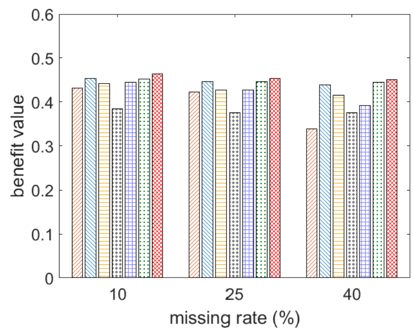

(b) Seeds data with $\phi = 2$.



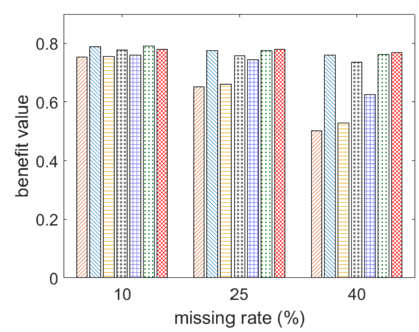

(c) Statlog data with $\phi = 6$.



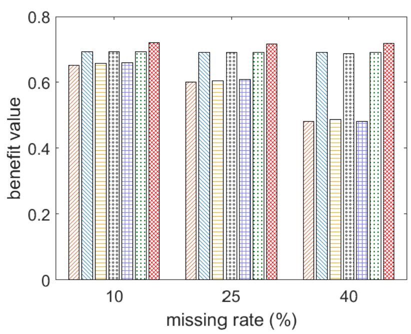

(d) Wine data with $\phi = 10$.



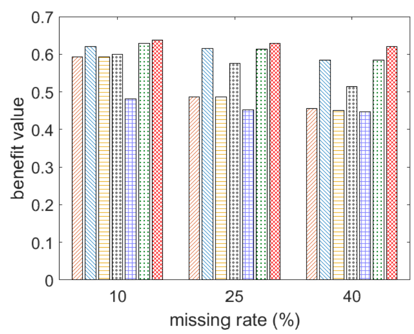

(e) Vehicle data with $\phi = 10$.



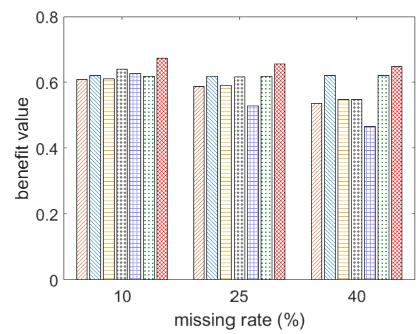

(f) Forest type data with $\phi = 20$.



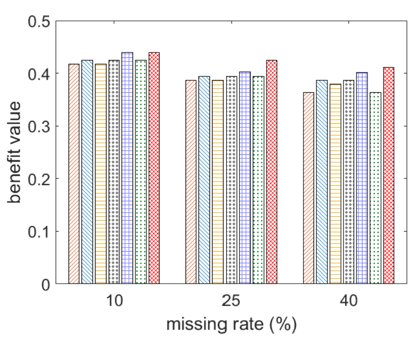

(g) Satimage data with $\phi = 20$.



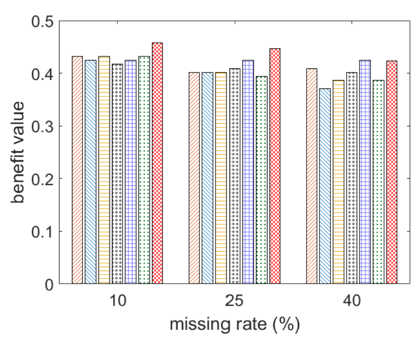

(h) Segment data with $\phi = 10$.



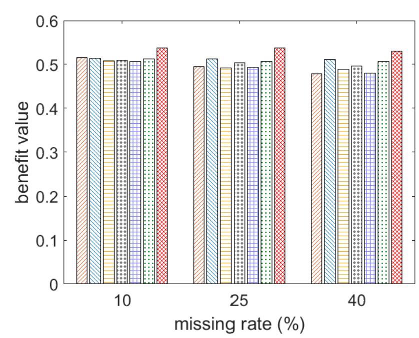

(i) Texture data with $\phi = 10$.




(j) Hayes data with $\phi = 2$.




(k) Knowledge data with $\phi = 2$.




(l) Abalone data with $\phi = 4$.


Figure 4: The benefit values of the used datasets with different missing rates.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3049511, IEEE Transactions on Knowledge and Data Engineering



**E-mail:** zhangzuowei0720@gmail.com

Zuo-wei Zhang was born in Anhui, China, in 1988. He received the master's degrees from Xi'an University of Architecture and Technology (XAUAT), China, in 2013, 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 belief function theory and its application in pattern recognition and multi-source information fusion.

**E-mail:** 1428375064@qq.com

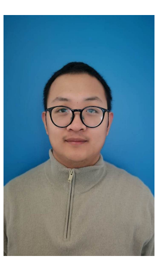
Zhe Liu was born in Shaanxi, China, in 1996. He received the bachelor's degrees from Xi'an University of Architecture and Technology Huaqing College, China, in 2018, and is currently pursuing the master's degree from Xi'an University of Architecture and Technology (XAUAT), China.

His current research interests include pattern recognition.

**E-mail:** Mazf@xauat.edu.cn

Zong-fang Ma was born in Anhui, China, in 1980. He received the bachelor's and master's degrees from Xi'an University of Architecture and Technology (XAUAT), China, and Ph.D. degrees from Northwestern Polytechnical University (NPU), Xi'an, China, in 2002, 2006 and 2011, respectively.

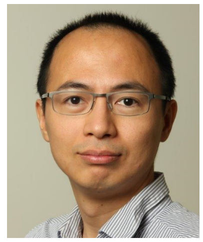
He is an Associate Professor with the School of Information and Control Engineering, XAUAT. His research interests include machine vision and pattern recognition.



**E-mail:** zyrbruce@gmail.com

Yi-ru Zhang is a research and teaching assistant at University of Rennes 1 in the DRUID team of IRISA laboratory. He received bachelor¡¯s degree from Huazhong University of Science and Technology (HUST), China, the engineer and master degrees from University of Pierre and Marie Curie (UPMC), France, and Ph.D degree from University of Rennes 1, France, in 2013, 2015 and 2020 respectively.

His current research interests cover the imperfect information reasoning with the theory of belief function, decision theory, preference learning, and pattern recognition.



**E-mail:** hawa@ntnu.no

[M'07] Hao Wang is an Associate Professor in the Department of Computer Science in Norwegian University of Science and Technology, Norway. He has a Ph.D. degree (2006) and a B.Eng. degree (2000), both in computer science and engineering, from South China University of Technology, China. His research interests include big data analytics, industrial internet of things, high performance computing, and safety-critical systems. He served as a TPC co-chair for IEEE CPSCom 2020, IEEE CIT 2017, ES 2017, IEEE DataCom 2015, and a senior TPC member for CIKM 2019. He has received prestigious awards such as IEEE outstanding paper award and IEEE outstanding leadership award. He is the Chair for Sub-TC on Healthcare of IEEE Industrial Electronics Society Technical Committee on Industrial Informatics.