BSC: Belief Shift Clustering

Zuo-Wei Zhang, Zhun-Ga Liu, Arnaud Martin, Kuang Zhou

BSC: Belief Shift Clustering

Zuo-wei Zhang, Zhun-ga Liu, Arnaud Martin, Kuang Zhou

***Abstract*—It is still a challenging problem to characterize uncertainty and imprecision between specific (singleton) clusters with arbitrary shapes and sizes. In order to solve such a problem, we propose a belief shift clustering (BSC) method for dealing with object data. The BSC method is considered as the evidential version of mean shift or mode seeking under the theory of belief functions. First, a new notion, called belief shift, is provided to preliminarily assign each query object as the noise, precise, or imprecise one. Second, a new evidential clustering rule is designed to partial credal redistribution for each imprecise object. To avoid the "uniform effect" and useless calculations, a specific dynamic framework with simulated cluster centers is established to reassign each imprecise object to a singleton cluster or related meta-cluster. Once an object is assigned to a meta-cluster, this object may be in the overlapping or intermediate areas of different singleton clusters. Consequently, the BSC can reasonably characterize the uncertainty and imprecision between singleton clusters. The effectiveness has been verified on several artificial, natural, and image segmentation/classification datasets by comparison with other related methods.**

***Index Terms*—Belief shift, mean shift, evidential clustering, uncertainty, mode seeking.**

I. Introduction

CLUSTERING analysis has been widely used in various fields [1]- [3], aiming to assign the dataset to different clusters where the objects are similar in one cluster but dissimilar to the objects in other clusters. A number of clustering techniques based on different philosophies have emerged.

K-means [4], as a representative of partition-based methods, provides a hard partition for the object depending on its distance to different cluster centers. The best cluster centers candidates are found by optimizing an objective function, typically the sum of the distance to a set of putative cluster centers. Motivated by [4], the early fuzzy c-means (FCM) [5] assigns the object to different clusters with various levels of support, which has robust characteristics for ambiguity. The extensions to the fuzzy-based method are widely studied in [6]- [13]. For example, the literatures [6]- [9] are dedicated to improving the performance and reassigning levels of support reasonably for objects. The literatures [10]- [12] aim to combine fuzzy-based methods and transfer learning. In [12], based on some useful knowledge available in the related scenes/domains, a set of transfer prototype-based fuzzy clustering methods is developed by exploiting the idea of leveraging knowledge from the source domain. The literature [13] develops a noise clustering method to make fuzzy-based methods more robust against noise. However, these fuzzy-based methods depending on symmetric distance [14], *e.g.* Euclidean distance, are not suitable for the clusters with arbitrary shapes and sizes in the space, such as nonspherical clusters [15] or imbalanced clusters [1], [4]. A few improved methods have been developed [1], [16]- [17]. For example, the literature [1] presents a multi-center (MC) clustering method to avoid the "uniform effect" of imbalanced data[1]. In MC, multiple centers (clusters) are used to represent each class. But it doesn't apply to close clusters and cannot be generalized to the clusters with arbitrary shapes and sizes.

Interestingly, density-based clustering methods can solve the above problems well, such as mean shift or mode seeking (MS) [19]- [20]. In MS, a probability density function is developed to represent the mode (cluster), then estimated by gradient ascent. Motivated by [19]- [20], some improved versions have been proposed [21]- [25]. For example, the literature [24] presents a semi-supervised framework for kernel mean shift clustering (SKMS), where the pairwise constraints are used to guide the clustering procedure. The literature [25] studies the properties of mean shift (MS)-type methods for estimating modes of probability density functions, and it finds several new properties of mode and corresponding density estimate sequences based on the function. Besides, the literature [26]- [27] replaces the kernel density estimate by K-nearest neighbor (KNN) ones. In KNN-based methods, the bandwidth is adaptively adjusted to make modes converge quickly. The area with low density has a large bandwidth, while a small bandwidth has a high density.

In addition, some other density-based methods have also been developed [15], [28]. For example, a new density peaks clustering (DPC) is proposed in [15]. In DPC, the centers are characterized by a higher density than neighbors and far away from other objects with higher densities. A density-based spatial clustering of applications with noise (DBSCAN) method is introduced in [28]. The objects in areas with densities lower than the density threshold are discarded as noise, and others with high density are assigned to different specific clusters.

Manuscript received February 3, 2021; revised August 24, 2021, January 25, 2022 and June 28, 2022; accepted September 3, 2022. Date of publication XXX; date of current version XXX. This work was supported in part by the National Natural Science Foundation of China under Grant U20B2067, Grant 61790552, and Grant 61790554; and in part by the Aeronautical Science Foundation of China under Grant 201920007001, and Grant 20182053023. This manuscript was recommended by Associate Editor XXX. *(Corresponding author: Zhun-ga Liu.)*

Zuo-wei Zhang is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China, and also with Univ Rennes, CNRS, IRISA, DRUID, 22300 Lannion, France (e-mail: zhangzuowei0720@gmail.com).

Zhun-ga is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: liuzhunga@nwpu.edu.cn).

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

Kuang Zhou is with the School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: kzhoumath@nwpu.edu.cn).

This article has supplementary material provided by the authors and color versions of one or more figures available at XXX.

Digital Object Identifier XXX.

[1]The "uniform effect" refers to that a part of objects belonging to majority classes are assigned to minority classes, which makes clusters have relatively uniform sizes, although input data have varied cluster sizes [1], [18].

Besides, the literature [29] successfully applies DBSCAN to large-scale datasets based on KNNs [30]. In addition, some spectral-based clustering methods [31], [32] also have shown good performance on arbitrarily shaped data. For example, an ultra-scalable ensemble clustering (U-SENC) method is proposed in [32], and it can effectively cluster extremely large-scale datasets with limited resources.

Some above methods have achieved good results for clusters with arbitrary shapes and sizes. However, there are some uncertain and imprecise information between these clusters. Here we argue that uncertainty is caused by insufficient knowledge; In contrast, imprecision is caused by fuzziness of knowledge. In clustering tasks, for example, one object simultaneously close to several singleton (specific) clusters is difficult to classify correctly since these close singleton clusters appear not distinguishable for this object. In this paper, we allow this object belonging to any singleton cluster or the unions of these singleton clusters (called meta-cluster) with different support degrees to characterize uncertainty and imprecision in the results. For example, if an object is assigned to a meta-cluster, we can say it is an *imprecise object*; Otherwise, it is called a *precise object* if it is only close to one singleton cluster. As shown in Fig. 1, there is a 7-class problem with two dimensions. The objects, distributed in the overlapping areas (*e.g.* Area 1) or intermediate areas (*e.g.* Area 4) of different clusters, are challenging to be identified by singleton clusters. If they are assigned forcibly, it may increase the risk of errors. Thus, it is unreasonable to assign these objects to singleton clusters only depending on current knowledge.

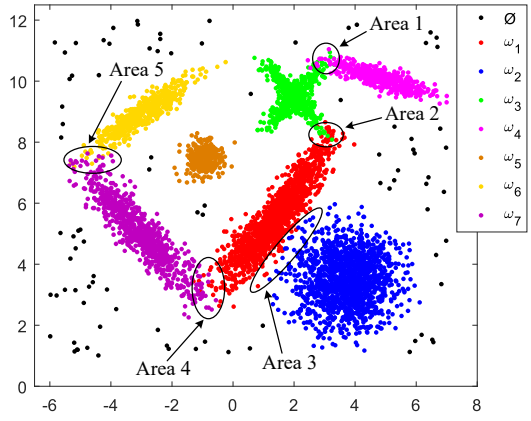


Fig. 1. Illustration of the noise and uncertainty between different clusters with arbitrary shapes and sizes.

The theory of belief functions (TBF) [33]- [36], also known as Dempster-Shafer theory, is appealing to deal with such uncertain and imprecise information. In TBF-based clustering analysis [2], [37]- [42], the object can be assigned to three kinds of clusters: singleton (specific) cluster (*e.g.* $\{\omega_i\}$), meta-cluster (*e.g.* $\{\omega_k, ..., \omega_t\}$) and the noise cluster represented by $\emptyset$. The "meta-cluster" is regarded as a new cluster and considered as a transition cluster among these different close singleton clusters [38]. For example, a new belief-peaks evidential clustering (BPEC) method is proposed in [2], which aims to overcome the shortages of density peaks clustering (DPC) [15] in characterizing uncertainty and imprecision in results. In BPEC, the cluster centers are first obtained based on the concept of "belief peak", and then each object is assigned to the nearest center under the TBF framework. However, this partition-based assignment mechanism is only used for spherical data. Therefore, it cannot effectively detect non-spherical data or generates backward results sometimes, a regression with respect to DPC.

In this paper, we propose a belief shift clustering (BSC) method, aiming to characterize uncertainty and imprecision between arbitrary clusters and thereby obtain a more robust and reasonable performance. The proposed BSC contains two characteristics: 1) Belief shift for preliminary credal partition; 2) Evidential clustering rule for partial credal redistribution. The main contributions can be summarized as follows.

1) A new notion of "belief shift" is proposed based on mean shift and the TBF, which can detect clusters with arbitrary shapes and sizes. In the process, each object is preliminarily assigned as the noise, precise object, or imprecise one.

2) An evidential clustering rule is designed to reassign the imprecise objects by partial credal redistribution, which provides a specific dynamic sub-framework with corresponding simulated centers to avoid the "uniform effect".

3) The proposed BSC can characterize uncertainty and imprecision between clusters with arbitrary shapes and sizes in results. Numerous experimental studies demonstrate the effectiveness of this argument.

The rest of this paper is organized as follows. After a brief introduction of belief functions in Section II, the belief shift clustering (BSC) is introduced in detail in Section III. The BSC is then tested in Section IV and compared with several typical methods. Section V includes some discussions such as the parameters, comparison of different methods, problems in applications, followed by conclusions.

II. Basics of Belief Functions

The theory of belief functions (TBF), also called Dempster-Shafer theory or evidence reasoning, which is a theoretical framework for reasoning with uncertain and imprecise information and has been widely used in various fields including data clustering [37]- [42], classification [43]- [46], and decision-making [47]- [49]. In the TBF, the frame of discernment $\Omega = \{\omega_1, ..., \omega_c\}$ is extended to the power-set 2^Ω, which contains all subsets of Ω. For example, if $\Omega = \{\omega_1, \omega_2, \omega_3\}$, then $2^\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$.

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

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

where A is called the focal element of $m(A)$ if $A \in 2^\Omega$ and $m(A) > 0$, and the set of all its focal elements is called the core of $m(A)$. For clustering tasks, these focal elements are also called clusters or categories. The TBF can generate three clusters: singleton cluster, meta-cluster, and noise cluster. Given $\Omega = \{\omega_1, ..., \omega_c\}$, we define these clusters as follows.

Definition 1. Singleton cluster, also called specific cluster and formed by A_j, aims to characterize the exact information, and has the following form: $A_j = \{\omega_k\},\ k \in [1, c]$.

Definition 2. Given a meta-cluster A_j, it is defined by the union (disjunction) of several singleton clusters and has the following form: $A_j = \{\omega_k, ..., \omega_l\}, k \neq l \in [1, c]$.

Definition 3. Noise cluster, represented by $\emptyset$, is considered a separate cluster and is defined as the set of those objects that are far from all singleton clusters. Once objects are assigned to the noise cluster, they are regarded as noise or outliers.

In TBF-based clustering, we usually consider that mass functions can characterize uncertainty, and meta-clusters can characterize imprecision in results. Moreover, we also review some TBF-based clustering methods in *Sections A, B, C of the supplementary file.*

The lower and upper bounds of probability associated with BBAs correspond to the belief function $Bel(.)$ and the plausibility function $Pl(.)$ with $\forall A \subseteq \Omega$ defined by:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{3}$$

$Bel(.)$ and $Pl(.)$ can be used for decision-making support when adopting pessimistic or optimistic attitudes. Similar to A in Eq. (1), both B in Eqs. (2) (3) and the subsequent C in Eqs. (4) (5) denote focal elements, *i.e.* $A, B, C \in 2^{\Omega}$.

The combination of mass functions plays a critical role in the TBF, where Dempster's (DS) rule has been widely used in the combination because it is commutative and associative. Assuming that $m_1(B)$ and $m_2(C)$ are two mass functions, then the combination with DS rule is defined by:

$$m(A) = \begin{cases} \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \mathcal{K}}, & \forall A \in 2^{\Omega}, A \neq \emptyset \\ 0, & A = \emptyset \end{cases} \tag{4}$$

where

$$\mathcal{K} = \sum\nolimits_{B \cap C = \emptyset} m_1(B) m_2(C),\ \mathcal{K} \neq 1. \tag{5}$$

The DS rule is applicable only when the denominator is strictly positive. Hence, the value of $\mathcal{K}$ must be smaller than 1 as $\mathcal{K} = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) < 1$.

In addition, we also review some basics of mean shift in *Sections D of the supplementary file.*

III. Belief shift clustering

In this section, the proposed belief shift clustering (BSC) method is presented in detail to characterize uncertainty and imprecision between clusters with arbitrary shapes and sizes in the space. The BSC mainly includes two parts:

1) Belief shift for preliminary credal partition. It mainly assigns each object in the query set as the noise (outlier) that is too far away from other singleton clusters, or a precise object with exact cluster information, or an imprecise object with several possible clusters information;

2) Evidential clustering rule for partial credal redistribution. It can further reassign each imprecise object obtained in the first part to different singleton clusters or meta-clusters including several singleton clusters that the object likely belongs to.

A. Belief shift for preliminary credal partition

Let us consider a dataset $\mathcal{X}$ including n objects with p attributes on the frame of discernment $\Omega = \{\omega_1, ..., \omega_c\}$. Belief shift for preliminary credal partition can be detailed as follows. For a specific object $\mathbf{x}_i$ $(i = 1, ..., n)$, we can obtain the neighbors by calculating the Euclidean distance between $\mathbf{x}_i$ and the others, defined by:

$$\|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{\sum_{q=1}^{p} (\mathbf{x}_{iq} - \mathbf{x}_{jq})^2} \tag{6}$$

where $\mathbf{x}_j \in \mathcal{X}$ and $\mathbf{x}_j \neq \mathbf{x}_i$. As a result, we can obtain the $\mathcal{K}_1$ neighbors, named $\mathbf{y}_1, ..., \mathbf{y}_k, ..., \mathbf{y}_{\mathcal{K}_1}$, with the corresponding minimum distance. These neighbors are selected from the whole dataset and each neighbor provides a piece of evidence represented by a mass function $m_{ik}(.)$ for the object $\mathbf{x}_i$ being a cluster center in a new frame of discernment $\Theta = \{\mathcal{C}, \mathcal{U}\}$. It is defined to describe the belief degree of the object as the cluster center ($\mathcal{C}$) or unknown ($\mathcal{U}$). A detailed description of this definition is included in *Section L of the supplementary file*. We assume that cluster centers are the objects with the highest global (local) density. If an object is close to all the neighbors, it has the potential to become a cluster center. Therefore, the mass function $m_{ik}(.)$ on Θ is defined by:

$$m_{ik}(E) = \begin{cases} \frac{1}{\mathcal{K}_1} \cdot e^{-\|\mathbf{x}_i - \mathbf{y}_k\|^2}, & \text{if } E = \mathcal{C} \\ 1 - \frac{1}{\mathcal{K}_1} \cdot e^{-\|\mathbf{x}_i - \mathbf{y}_k\|^2}, & \text{if } E = \mathcal{U} \end{cases} \tag{7}$$

where $||.||$ is the Euclidean distance. In Eq. (7), if $E = \mathcal{C}$, for example, it means that the object $\mathbf{x}_i$ is a cluster center. The mass function $m_{ik}(\mathcal{C})$ can be regarded as the support degree that the object $\mathbf{y}_k$ believes the object $\mathbf{x}_i$ as a cluster center. We can see from Eq. (7) that the nearer the neighbor to $\mathbf{x}_i$, the larger the $m_{ik}(\mathcal{C})$ obtained, *i.e.* this neighbor strongly supports the object $\mathbf{x}_i$ as a cluster center.

Afterward, $\mathcal{K}_1$ pieces of evidence provided by the neighbors can be fused by the DS rule, thereby we can obtain the degree of belief $m_i(.)$ on $\tilde{\Omega}$ that the object $\mathbf{x}_i$ is a cluster center or not, defined by:

$$m_i(E) = \bigoplus_{k \in [1, \mathcal{K}_1]} m_{ik}(E) \tag{8}$$

where $\bigoplus$ represents the DS rule operation symbol. We can get $m_i(E) = Bel_i(E)$ by deriving Eq. (8). The proof process is included in *Section L of the supplementary file*. Thus, we can redefine the degree of belief $Bel_i(E)$ as follows:

$$Bel_i(E) = \begin{cases} 1 - \prod_{k=1}^{\mathcal{K}_1} (1 - m_{ik}(\mathcal{C}))\ , & \text{if } E = \mathcal{C} \\ \prod_{k=1}^{\mathcal{K}_1} m_{ik}(\mathcal{U})\ , & \text{if } E = \mathcal{U}. \end{cases} \tag{9}$$

By doing so, we further calculate the belief degrees of all the n objects. In the process, each object is regarded as the initial cluster center for belief shift to eliminate the negative impact

of random selection[2]. Thus, the centers are all real objects. In the process, for the specific object $\mathbf{x}_i$ $(i = 1, ..., n)$, as the initial cluster center, it will shift to the new center (object), named $\mathbf{x}_\mu$, corresponding to the neighbor $\mathbf{y}_k$ with the highest belief degree and defined by:

$$\mathbf{x}_\mu = \arg\max_{\mathbf{y}_k \in \mathcal{X}} \{Bel_{i1}(\mathcal{C}), ..., Bel_{i\mathcal{K}_1}(\mathcal{C})\} \tag{10}$$

where $Bel_{ik}(\mathcal{C})$ $(k = 1, ..., \mathcal{K}_1)$ represents the belief degree of the k-th neighbor $\mathbf{y}_k$ as a cluster center. Then, the center $\mathbf{x}_\mu$ keeps shifting until the belief degree $Bel_\mu(\mathcal{C})$ is higher than that of all the new neighbors and defined by:

$$Bel_\mu(\mathcal{C}) \geq \max\{Bel_{\mu 1}(\mathcal{C}), ..., Bel_{\mu\mathcal{K}_1}(\mathcal{C})\} \tag{11}$$

We can obtain the final cluster center $\mathbf{x}_\mu$ that $\mathbf{x}_i$ converges to. Similarly, all objects will converge to c cluster centers, where these centers have the highest belief degrees than their neighbors. Moreover, although the object $\mathbf{x}_i$ converges to a singleton cluster, as a neighbor of different query objects, it may also be searched by other singleton clusters. In this case, we temporarily consider it as an imprecise object. To show belief shift more intuitively and explain its difference from typical mean shift, we present the following example.

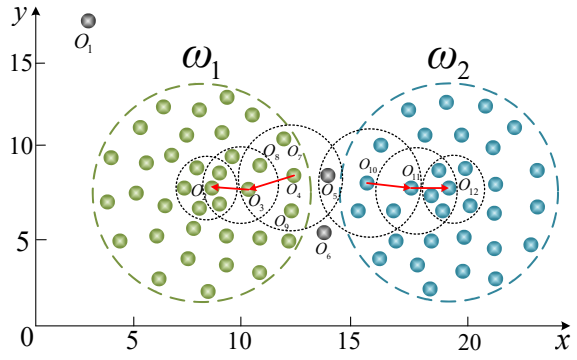


Fig. 2. Illustration of the belief shift process.

Fig. 2 shows a 2-class problem with two dimensions. We assume that the objects O_2 and O_{12} have the highest degree of beliefs as the cluster centers. In the process, $\mathcal{K}_1 = 5$ is the default for each query object. For more intuitively, the $\mathcal{K}_1$ neighbors are in the range of the black dotted line, and the trajectory of belief shift is marked as a red line with an arrow. For example, as the initial center, the object O_4 will find the $\mathcal{K}_1$ neighbors (*i.e.* O_3, O_5, O_7, O_8, O_9), and then shifts to the neighbor O_3 with the highest belief degree, and finally converges to the cluster center O_2, *i.e.* $O_4 \in \{\omega_1\}$. Similarly, the object O_{10} will converge to the cluster center O_{12}, *i.e.* $O_{10} \in \{\omega_2\}$. For objects like O_4 and O_{10}, both mean shift (MS) and belief shift will assign them to singleton clusters $\{\omega_1\}$ and $\{\omega_2\}$ directly because each object is searched by only one cluster. Hence, these objects are called *precise objects*. This is the same for MS and belief shift.

For the object O_5, however, MS will also assign it to the cluster $\{\omega_1\}$ or $\{\omega_2\}$ directly. The reason is that MS depends only on the times searched by these two clusters. Thus, the cluster with the most search times will have this object. However, it may have a huge error risk because the search times are related to the randomly initialized cluster centers in MS. Thus, it is unreasonable to depend only on the search times. Unlike MS, we consider that as a neighbor, once different clusters search the object O_5, it indicates that O_5 may be distributed in these clusters' overlapping or intermediate areas. We also have a similar inference from the object O_6 if interesting. Therefore, it is unreasonable to assign the object O_5 (or O_6) to the singleton cluster $\{\omega_1\}$ or $\{\omega_2\}$ since it will significantly increase the risk of errors. In this case, they are temporarily assigned to the specific edited framework $\mathcal{M}_i = \{\omega_1, \omega_2\}$, $\mathcal{M}_i \subseteq \Omega$ and $i = 5, 6$, as imprecise objects to wait for the next credal redistribution. In addition, although the object O_1 can converge to the cluster $\{\omega_1\}$, as a neighbor of other query objects, it will not be searched by different clusters because it is too far away from other objects. The object O_1 is more suitable for assigning to the noise cluster. That is, we prefer to consider the object O_1 as noise. However, MS cannot deal with these above cases.

By the above analysis, the query object $\mathbf{x}_i$ have two general indexes: 1) The number of different clusters that search this object, denoted as $|\mathcal{M}_i|$; 2) As a neighbor, the number of times this object is searched by other ones, recorded as $\mathcal{T}_i$. The total number $\mathcal{T}_i$ of searched times for $\mathbf{x}_i$ is defined by:

$$\mathcal{T}_i = \sum_{\{\omega_j\} \in \mathcal{M}_i} \mathcal{T}_i^j \tag{12}$$

where $\mathcal{T}_i^j$ represents the number of times that the object $\mathbf{x}_i$ is searched by the other ones that converge to the cluster $\{\omega_j\}$. If $\mathbf{x}_i$ is searched significantly less than that of others, it is assigned directly to the noise cluster, *i.e.* $\emptyset$, defined by:

$$\emptyset = \{\mathbf{x}_i \mid \mathcal{T}_i \leq \left[\overline{\mathcal{T}}^\alpha\right]\} \tag{13}$$

with

$$\overline{\mathcal{T}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}_i \tag{14}$$

where $[\cdot]$ is the rounding symbol, and $\overline{\mathcal{T}}$ is the mean of $\mathcal{T}_i$ for all objects. α is the outlier adjustment factor, which controls the number of objects assigned to the noise cluster, *i.e.* $\emptyset$.

The flowchart of preliminary adaptive credal partition by belief shift is given in *Section E of the supplementary file*. We can see that 1) the query object $\mathbf{x}_i$ is regarded as the noise (outlier) if it is searched as a neighbor by very few or even no other objects. This also means that $\mathbf{x}_i$ is far away from the others. 2) If $|\mathcal{M}_i| = 1$ with $\mathcal{T}_i > \left[\overline{\mathcal{T}}^\alpha\right]$, it indicates that $\mathbf{x}_i$ is a precise object with exact cluster information and should be assigned to the singleton cluster $\mathcal{M}_i$. 3) Otherwise, it is temporarily regarded as an imprecise object.

Here we need to clarify that belief shift has two functions. The first one is to obtain the c cluster centers. The second one is to assign each object as the noise, precise, or imprecise one. To avoid the "uniform effect" when reassigning imprecise objects, we generate new simulated cluster centers for each imprecise object in subsection III-B. In this case,

[2]The class with few objects is hard to be identified accurately if we just initial part objects as cluster centers like mean shift. Thus, BSC initials each object as a cluster center to implement belief shift.

the c cluster centers obtained in this part are not employed in subsection III-B. This also means getting the c cluster centers is not our goal. Therefore, we highlight the preliminary credal partition of each object by belief shift while weakening obtaining the cluster centers.

B. Evidential clustering rule for credal redistribution

For the imprecise object $\mathbf{x}_i$, the masses of belief will be partially redistributed by evidential clustering rule to reassign $\mathbf{x}_i$ to different clusters under the corresponding specific dynamic sub-framework $\mathcal{M}_i$ ($|\mathcal{M}_i| > 1$). After preliminary credal partition based on belief shift, we consider that there are n_1 $(0 < n_1 < n)$ imprecise objects in the set $\mathcal{X}_{im} = (\mathbf{x}_1, ..., \mathbf{x}_{n_1}) \in \mathbb{R}^{p\times n_1}$ and n_2 $(0 < n_2 < n)$ precise objects in the set $\mathcal{X}_{pr} = (\mathbf{x}_1, ..., \mathbf{x}_{n_2}) \in \mathbb{R}^{p\times n_2}$, $n_1 + n_2 = n$. For the imprecise object $\mathbf{x}_i \in \mathcal{X}_{im}$, it well be reassigned under the new edited dynamic framework $\mathcal{F}_i \subset 2^\Omega$ with $2^{|\mathcal{M}_i|} - 1$ elements under the TBF. The credal partition $m_{ij} \triangleq m_i(A_j) \in \mathbb{R}^{2^{|\mathcal{M}_i|}-1}$ with the j-th focal element A_j in $\mathcal{F}_i$, *i.e.* $A_j \in \mathcal{F}_i$, is provided for each imprecise object $\mathbf{x}_i$ $(i = 1, ..., n_1)$. For example, for the imprecise object $\mathbf{x}_i$, if $\mathcal{M}_i = \{\omega_1, \omega_3, \omega_5\}$ with $|\mathcal{M}_i| = 3$ after belief shift, then we have $\mathcal{F}_i = \{\{\omega_1\}, \{\omega_3\}, \{\omega_5\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_5\}, \{\omega_3, \omega_5\}, \mathcal{M}_i\}$.

In a credal partition [37]- [39], the dataset converges to c clusters and related meta-clusters by alternating iterations of the center matrix and the mass of belief matrix. Although the real (final) cluster centers have been obtained by belief shift, we cannot use them to iterate the masses of belief directly since it is unreasonable to use one fixed center to represent the class with arbitrary shape and size. For example, when dealing with imbalanced data, these centers tend to assign the objects in the majority classes to the minority classes [1]. Here we provide specific simulation centers for related clusters supervised by $\mathcal{F}_i$ when reassigning each imprecise object $\mathbf{x}_i$. We can further explain this by an example, as shown in Fig. 3, with a 2-class dataset $\Omega = \{\omega_1, \omega_2\}$.

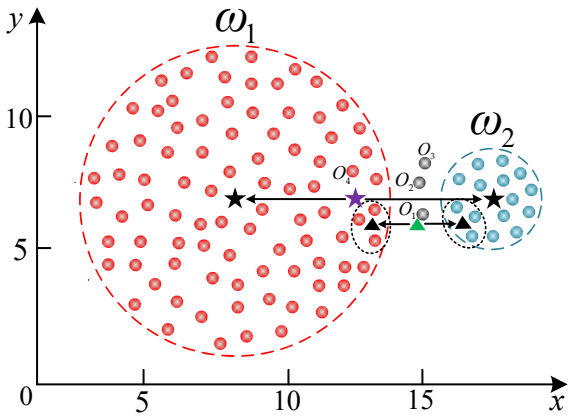


Fig. 3. Illustration of the simulated cluster centers based on imbalanced data.

In Fig. 3, the black pentagrams represent the cluster centers of $\{\omega_1\}$ and $\{\omega_2\}$ after belief shift, and the objects specifically assigned to different clusters are precise objects while the objects O_1, O_2 and O_3 are imprecise objects. We can see that the center of meta-cluster $\{\omega_1, \omega_2\}$ marked as purple pentagram is obtained by calculating the mean of $\{\omega_1\}$ and $\{\omega_2\}$ in a credal partition. However, the center is located in the cluster $\{\omega_1\}$, which is obviously unreasonable and runs counter to our intuitive perception. In this case, if we use the unreasonable meta-cluster center directly, those objects like O_4 that originally belong to the cluster $\{\omega_1\}$ will be assigned to the meta-cluster $\{\omega_1, \omega_2\}$ and the objects like O_1 will be assigned to the cluster $\{\omega_2\}$. In fact, the meta-cluster center should be located in the midpoint of the edges.

To address such issues, the simulated cluster centers based on KNNs are calculated to assign each imprecise object partially. Here the neighbors are defined as $\mathcal{K}_2$ objects, $\mathbf{y}_k$, $k = 1, ..., \mathcal{K}_2$, based on Eq. (6). That is, we should find $\mathcal{K}_2$ neighbors from each singleton cluster included in $\mathcal{M}_i$ to simulate the corresponding center. Then, the simulated singleton cluster centers are employed to calculate the related meta-cluster centers. That is, we adaptively provide specific cluster centers included in $\mathcal{F}_i$ for each imprecise object $\mathbf{x}_i$. For example, the imprecise object O_1 finds different neighbors included in the black dashed circle in clusters $\{\omega_1\}$ and $\{\omega_2\}$ respectively, as shown in Fig. 3. The black triangle represents the simulated center obtained by the mean of neighbors, and the midpoint marked as the green triangle is the meta-cluster center for O_1. We can intuitively see that the meta-cluster center obtained by simulated centers is located at the halfway of $\{\omega_1\}$ and $\{\omega_2\}$, which is more reasonable than the center marked as the purple pentagram. Based on these, the singleton cluster center $\mathbf{v}_{il}$ is defined by:

$$\mathbf{v}_{il} = \frac{1}{\mathcal{K}_2}\sum_{k=1}^{\mathcal{K}_2}\mathbf{y}_k^l,\ l = 1, ..., c \tag{15}$$

where $\mathbf{y}_k^l$ $(k = 1, ..., \mathcal{K}_2)$ represents the k-th neighbor in the cluster $\{\omega_l\} \subset \mathcal{X}_{pr}$. Besides, the meta-cluster center is the mean of that of the related singleton cluster centers.

As a result, based on the simulated center matrix V, the evidential clustering rule to update the mass of belief for the object $\mathbf{x}_i$ is defined by:

$$m_{ij} = \frac{D_{ij}^{-2/(\beta-1)}}{\sum\limits_{j|A_j\in F_i} D_{ij}^{-2/(\beta-1)}} \tag{16}$$

subject to

$$\mathcal{J}_{BSC}(M, V) = \sum_{i=1}^{n_1}\sum_{j|A_j\in\mathcal{F}_i} m_{ij}^\beta D_{ij}^2 \tag{17}$$

and

$$D_{ij}^2 = \begin{cases} d_{ij}^2, & \text{if } |A_j| = 1 \\ \frac{\sum\limits_{\omega_l\in A_j}\gamma^{-1}d_{il}^2 + d_{ij}^2}{|A_j|+1}, & \text{if } |A_j| > 1 \end{cases} \tag{18}$$

where d_{ij} is the Euclidean distance between $\mathbf{x}_i$ and the center of cluster A_j; d_{il} represents the distance from $\mathbf{x}_i$ to the singleton cluster centers in A_j such that $|A_j| > 1$. n_1 is the number of imprecise objects. The tuning parameter β, such that $\beta > 1$, is a weighting exponent [37] and γ is the threshold to control the number of objects in meta-

clusters [38]. We can find that the distance between the object $\mathbf{x}_i$ and the meta-cluster A_j with $|A_j| > 1$ depends not only on the distance from $\mathbf{x}_i$ to the meta-cluster center, but also on the distance between $\mathbf{x}_i$ and the centers of all the singleton clusters included in A_j. If interesting, please refers to Eq. (11) in *Section C of the supplementary file*.

Different from CCM [38], however, we do not consider noise clustering here because we have solved this problem well in the preliminary credal partition. Since the simulated cluster centers are reliable, only one update is needed to produce the masses of belief that each imprecise object $\mathbf{x}_i$ belongs to different clusters. This can decrease the computation brought by the iterative process while ensuring rationality. By doing so, each imprecise object in the set $\mathcal{X}_{im}$ is credal redistributed again based on the evidential clustering rule. Part of objects is reassigned to singleton clusters in the process, which means that these objects are precise ones with exact cluster information. By contrast, the rest are reassigned to related meta-clusters to characterize imprecision caused by fuzziness of knowledge. This prudent decision-making can characterize the uncertainty and imprecision between clusters with arbitrary shapes and sizes, which may be critical in some applications. By the way, the pseudo-code is presented in **Algorithm 1** to show how BSC works and illustrate its basic principle clearly.

Algorithm 1 Belief shift clustering

Require: Dataset: $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$; Given the parameters: $\mathcal{K}_1, \mathcal{K}_2, \alpha, \beta, \gamma$.
Ensure: Cluster decision results.
Step 1
Search the neighbors for all objects using Eq. (8);
Calculate $(Bel_i(\mathcal{C}))$ for all objects using Eqs. (9)-(11);
for i = 1 to n
repeat
Each object is employed to belief shift using Eq. (12);
until Satisfy the judgment condition of Eq. (13).
end
Assign the outlier using Eqs. (14)-(16);
Assign precise and imprecise objects using $|\mathcal{M}_i|$ and $\mathcal{T}_i$;
Step 2
for i = 1 to n_1
Calculate simulated cluster centers using Eq. (17);
Calculate the meta-cluster centers based on the simulated singleton cluster centers;
Reassign each imprecise object again using Eq. (18).
end
Return: Output the results.

C. Tuning of parameters

In BSC, some parameters including $\alpha, \gamma, \beta, \mathcal{K}_1, \mathcal{K}_2$ play a very important role, and they should be selected in advance to implement the proposed BSC method. α is the outlier adjustment factor, which controls the number of objects regarded as the noise. In applications, α is tuned accordig to the used datasets. In general, the bigger α causes the more objects assigned to the noise cluster, and we recommend that $\alpha \in [0, 0.6]$ and take $\alpha = 0.3$ as the default. The weighting factor γ can be used to control the number of objects in meta-clusters. The smaller γ is, the fewer objects are assigned to meta-clusters, which will increase the number of misclassified objects. Whereas, γ is not the bigger the better since a big value of γ will lead to high imprecision. Therefore, the selection of γ should be based on the imprecision rate that one can accept. Some works (*e.g.* CCM [38]) have discussed the influence of this parameter on clustering results and provided the ranges in applications. Similar to CCM, we recommend $\gamma \in [0.5, 2.5]$ and generally take $\gamma = 1$ as the default. The tuning parameter β is a weighting exponent. Similar to FCM [5] and ECM [37], $\beta = 2$ is used as the default. $\mathcal{K}_1$ is the number of not only the neighbors that are used to provide the pieces of evidence for the object being a cluster, but also the neighbors that the objects are looking for in the process of belief shift. The value of $\mathcal{K}_1$ should not be too small since it may cause the object to fall into the local maximum belief degree during belief shift. Whereas some clusters with very close data distribution may not be able to correctly distinguish if too large $\mathcal{K}_1$ value is set. Here the value of $\mathcal{K}_1$ is determined by the number n of objects in the dataset, and we find that $\mathcal{K}_1 \in [0.01n, 0.2n]$ can be used as the default in most cases according to numerous experiments. $\mathcal{K}_2$ is the number of neighbors that are used to get simulated cluster centers of different clusters, and it does not need to take too large. Thus, we recommend a common default value, *i.e.* $\mathcal{K}_2 = 7$. Both $\mathcal{K}_1$ and $\mathcal{K}_2$ are the open parameters, and we try to optimize them based on other technologies in the future.

IV. Experiment applications

We have done four experiments to evaluate the performance of the proposed BSC method with respect to K-means [4], DBSCAN [28], MC [1], DPC [15], MS-type [25], U-SENC [32] and BPEC [2]. Experiment 1 and Experiment 2, based on particular synthetic datasets, are used to illustrate the use of BSC and the limitations of other methods. Experiment 3 with real images is presented to evaluate the effectiveness of the proposed BSC method in image segmentation. Experiment 4 is used to reveal the potential of BSC in image classification with face datasets. All parameters are defaults except the ones we adjust. The detail is presented in *Section F of the supplementary file.*

The error rate and impression rate are used as performance indexes of different methods [38]- [39]. The error rate, denoted as R_e (In %), is calculated by $R_e = T_e/T$. T_e is the number of clustering errors, and T is the number of objects under test. The imprecision rate, denoted as R_i (In %), is calculated by $R_i = T_i/T$. T_i is number of objects assigned to the meta-clusters. Additionally, the Credal Rand Index (CRI) [52] based on TBF, regarded as the evidential version of the widely used performance index called Adjusted Rand Index (ARI) [53], is employed to measure the closeness of the credal partition and the truth. In general, CRI provides the overall

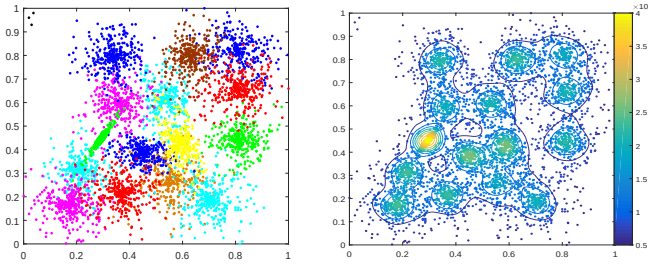
performance of different methods when clustering different datasets, defined by:

$$CRI = (M, M^*) = \frac{\sum_{i<j} pl_{ij}(s)^{r^*_{ij}} pl_{ij}(\neg s)^{1-r^*_{ij}}}{n(n-1)/2} \quad (19)$$

where M represents the evidential partition, and M^* is the true hard partition. n is the number of objects in the dataset. $r^*_{ij} = 1$ if the i-th and j-th object truly belong to the same cluster, and $r^*_{ij} = 0$, otherwise. Particularly, CRI and ARI are equal when comparing the closeness of the hard partition to the truth. In this paper, the performance index is uniformly denoted as "CRI". The higher the CRI value, the better the performance of the method. We define $\omega_i \triangleq \{\omega_i\}$ and $\omega_{k,...,t} \triangleq \{\omega_k, ..., \omega_t\}$ in the figures for notation conciseness.

A. Experiment 1

In this experiment, we employ a synthetic dataset named **SD15** to validate the effectiveness of BSC and reveal the limitations of hard partitions, including MC, DPC, and MS-type in clustering data with high overlap and noise.



(a) Original data. (b) Probability distribution.

Fig. 4. The basic information of the **SD15** dataset.

The **SD15** is from [54] and contains 5000 data points in 15 classes. In addition, we give 3 noisy points marked by black dots in the top left corner of **SD15**. The **SD15** is shown in Fig. 4(a) while Fig. 4(b) reveals the probability distribution. Various colors mark the data points included in different clusters. All the attributes are normalized into [0,1] by the min-max rule introduced in [2] to eliminate the influence of differences in various dimensions. Here, we set $Eps = 0.03$ and $MinPts = 40$ in DBSCAN. We choose $f = 0.02$ in DPC, $r = 0.08$ in MS-type, and $\mathcal{K}_1 = 80$ in BSC, respectively.

We can see from Fig. 4(a) that these 15 classes are partly overlapped on their borders, and the points in these areas are difficult to be classified. Figs. 5(a)-(f) show the clustering results of K-means, DBSCAN, MC, DPC, MS-type and U-SENC. We can see that the points in these overlapping areas are all assigned to singleton clusters by these methods and most of them are misclassified. It is worth noting that the noisy points marked with black dots are far from the other points, and they cannot be detected by K-means, MC, DPC, and U-SENC but assigned to singleton clusters. Although DBSCAN can distinguish the three noisy points, it assigns too many objects that initially belong to singleton clusters to the noise cluster. In particular, MS-type yields a singleton cluster for the noisy data, but it also does not assign these special points correctly due to the probability framework's limitations that do not introduce the noise cluster. Fig. 5(g) shows the trajectory of the belief shift in BSC. The trajectories of the objects are marked as red lines, and they finally coverage to the 15 points with the highest belief degrees marked by blue dots. Fig. 5(h) shows the preliminary credal partition of BSC for the points based on belief shift. The 3 noisy points are only searched by a few in the belief shift process since they are far from others. By contrast, as a neighbor, the points searched multiple times by others and searched by only one cluster can be directly assigned to specific clusters. Whereas the points, marked as gray and searched by multiple clusters in the process, are imprecise points, and most of them lie in overlapping areas of different clusters. The final clustering results of BSC are shown in Fig. 5(i). BSC provides masses of belief of the imprecise objects associated with singleton clusters to characterize the uncertainty in the results. Some of these objects are assigned to singleton clusters marked by points with different colors if they have the highest support degrees belonging to these clusters. By contrast, BSC assigns some imprecise objects lying in overlapping areas to proper meta-clusters marked by crosses with different colors. By doing so, BSC can effectively reduce the errors and characterize imprecision between different clusters. Therefore, BSC exhibits the lowest error rate and highest value of CRI concerning comparison methods.

B. Experiment 2

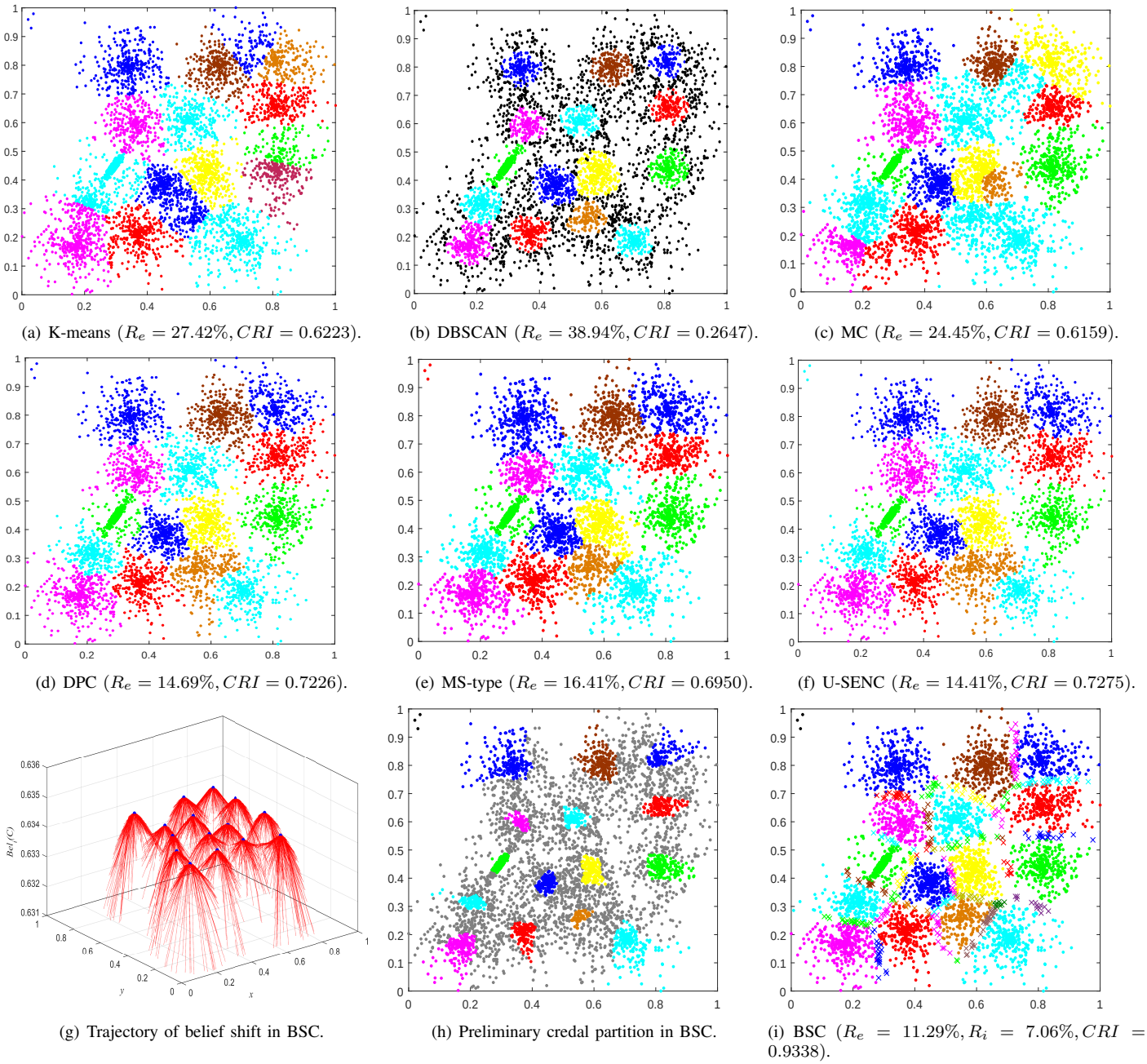
This experimentreveals the limitations of the typical BPEC method. Here we only show partial results. We investigate the performance of BSC and BPEC in clustering a 4-class dataset named **SD4** that contains not only overlapping areas in different clusters but also has arbitrary shapes and sizes in the space. The dataset consists of 3300 data points with two dimensions. The points arise from a mixture of four bivariate Gaussian densities are given in Table I, where μ_i is the means vector and Σ_i is the covariance matrices. N_i represents the number of data points in different classes.

Table I: The basic information of the **SD4** dataset

Data	Indexes	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_4\}$
SD4	μ_i	$[2.5, 5]^T$	$[7.5, 5]^T$	$[5, 7]^T$	$[5, 1.5]^T$
	Σ_i	$\begin{pmatrix} 1 & 0 \\ 0 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0 \\ 0 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
	N_i	500	500	300	2000

Fig. 6(a) intuitively shows the distribution of the points in this dataset, and the probability distribution is given by Fig. 6(b). Here $Eps = 0.35$ and $MinPts = 20$ are set in DBSCAN. We take $K = 70$, $\alpha = 3$, $\Delta = 5$ in BPEC and $\mathcal{K}_1 = 200$ in BSC, respectively.

Fig. 6(c), (f) shows the results of BPEC and BSC on the **SD4** dataset. They can carefully assign the points in overlapping areas of different classes to appropriate meta-clusters. Both of these methods can characterize the uncertainty and imprecision of these points and reduce the risk of errors. As a result, we can see that BSC yields a lower error rate and imprecision rate than BPEC. Although BPEC can



(a) K-means ($R_e = 27.42\%, CRI = 0.6223$). (b) DBSCAN ($R_e = 38.94\%, CRI = 0.2647$). (c) MC ($R_e = 24.45\%, CRI = 0.6159$).

(d) DPC ($R_e = 14.69\%, CRI = 0.7226$). (e) MS-type ($R_e = 16.41\%, CRI = 0.6950$). (f) U-SENC ($R_e = 14.41\%, CRI = 0.7275$).

(g) Trajectory of belief shift in BSC. (h) Preliminary credal partition in BSC. (i) BSC ($R_e = 11.29\%, R_i = 7.06\%, CRI = 0.9338$).

Fig. 5. The results of different methods on the **SD15** dataset.

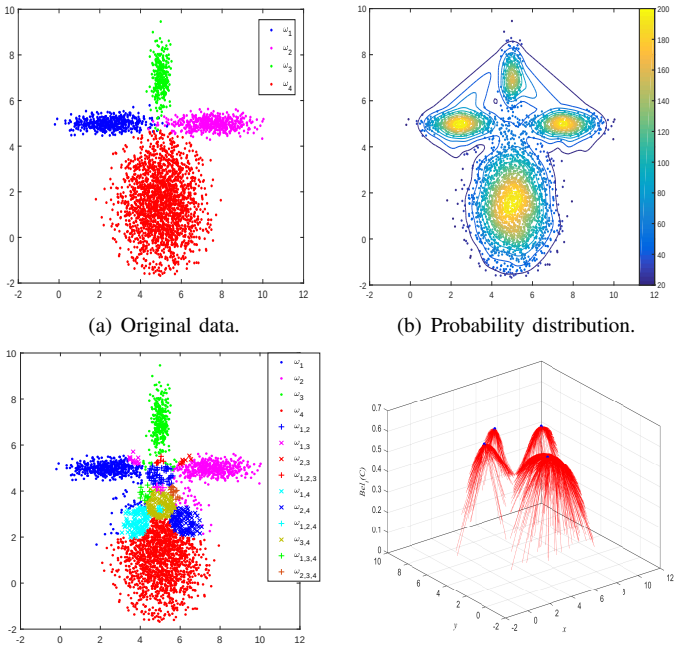
also provide the credal partition, it assigns a part of points belonging to the majority class to the minority classes in dealing with such a dataset. For example, some points in the class $\{\omega_4\}$ are assigned to $\{\omega_1\}$ and $\{\omega_2\}$, since the methods like BPEC based on symmetric distance only consider the symmetry of points in the space and don't take into account the distribution or the number of points in different classes. Thus, in this case, the center of meta-clusters tends to shift to the singleton class with majority points so that we may obtain unreasonable clustering results. By contrast, BSC simulates cluster centers using neighbors at the boundary of different classes, effectively decreasing the negative impact of clusters with arbitrary shapes and sizes. In addition, the results of other methods with experimental analysis are given in *Section G of the supplementary file*. In addition, due to space limitations, we will further discuss the differences and connections between BSC and existing typical TBF-based clustering methods in *Section L of the supplementary file*.

C. Experiment 3

In this experiment, a medical image and a natural image are employed to demonstrate the effectiveness of BSC in image segmentation. We obtain a true color dermoscopic image (invasive malignant melanoma) of 66 × 86 pixels, named **DI2**, from the EDRA Interactive Atlas of Dermoscopy [55], as shown in Fig. 7(a). It consists of two classes, including lesion

Table II: The results of different methods on the **DI2** and **NI2** datasets

Data	Indexes	K-means	DBSCAN	MC	DPC	MS-type	BPEC	U-SENC	BSC
DI2	$Re(\%)$	4.81	29.86	4.61	3.91	6.93	3.11	4.62	**2.24**
	$Ri(\%)$	/	/	/	/	/	3.46	/	**2.40**
	CRI	0.8080	-0.0087	0.8201	0.8554	0.7255	0.9031	0.8205	**0.9476**
NI2	$Re(\%)$	4.33	25.15	5.76	4.83	3.85	2.85	24.89	**2.63**
	$Ri(\%)$	/	/	/	/	/	3.50	/	**3.17**
	CRI	0.8156	-0.0399	0.7464	0.7878	0.8313	0.9106	0.2358	**0.9235**

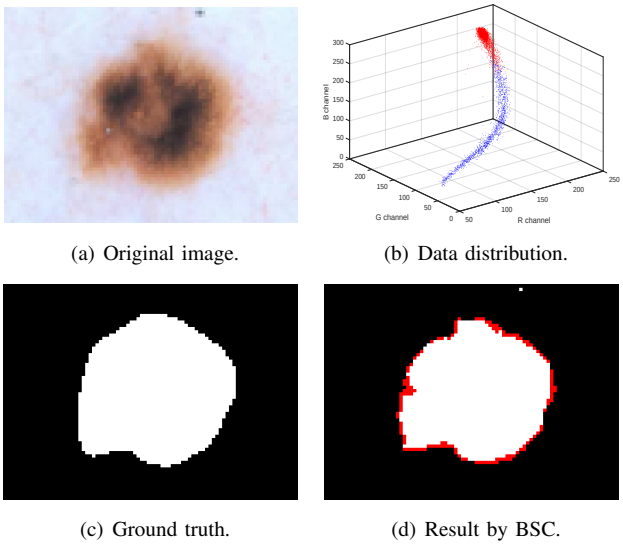


(a) Original data. (b) Probability distribution.

(c) BPEC (R_e = 3.79%, R_i = 14.55%, CRI = 0.8368). (d) Trajectory of belief shift in BSC.

(e) Preliminary credal partition in BSC. (f) BSC (R_e = 1.03%, R_i=2.12%, CRI = 0.9571).

Fig. 6. The results of BPEC and BSC on the **SD4** dataset.



(a) Original image. (b) Data distribution.

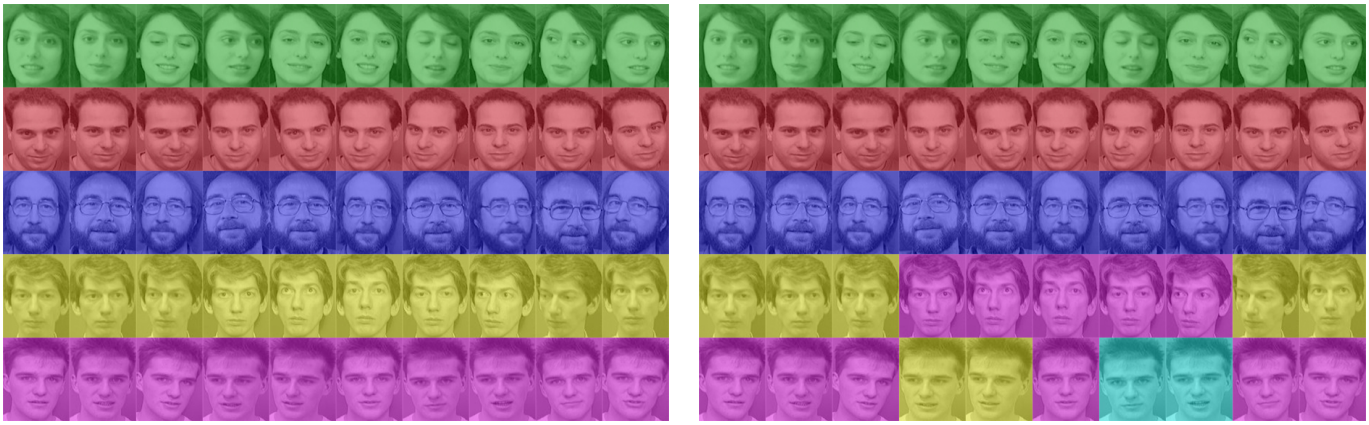
(c) Ground truth. (d) Result by BSC.

Fig. 7. The medical image and the result of BSC.

and non-lesion, and Fig. 7(c) shows the ground truth. The distribution of the pixels is shown in Fig. 7(b). The blue and red points with three dimensions, including R, G, and B value, represent the pixels of the lesion and non-lesion according to the ground truth, respectively. Here, we set $Eps = 0.05$ and $MinPts = 40$ in DBSCAN. We choose $f = 0.02$ in DPC and $r = 150$ in MS-type. $K = 2000$, $\alpha = 3$, $\Delta = 500$ and $\mathcal{K}_1 = 1000$ are set in BPEC and BSC, respectively.

We can see from Fig. 7(a) that the lesion edge is ambiguous and the distribution of the pixels given in Fig. 7(b). It intuitively reveals some pixels distributed in the overlapping area of different classes. These pixels correspond to the lesion edge, and they are challenging to be distinguished accurately. The segmentation results of comparison methods are shown in *Section H of the supplementary file*. We can see from the results of K-means, DBSCAN, DPC, MC, MS-type, and U-SENC that some pixels in the lesion edge are assigned to the specific cluster (*i.e.* lesion or non-lesion), and most of them are misclassification. The parameters have a significant impact on the performance of DBSCAN, and it has poor performance on this dataset with these parameters. Although MS-type is latter than MC and DPC, it is worse than them in this dataset since the performance is strongly affected by the initial cluster centers. Interestingly, we can also observe from Fig. 7(h),(d) that BPEC and BSC cautiously assign the pixels in lesion edge to the meta-cluster composed of the lesion and non-lesion under the TBF. That is, the pixels in this area are indistinguishable. However, the error rate and imprecision rate of BSC are lower than that of BPEC, and the overall performance of BSC is superior to comparison methods. The reason is that BSC has the advantage of characterizing uncertainty and imprecision between clusters with arbitrary shapes and sizes under the TBF.

Doing so can reduce the risk of error and characterize the imprecision between different clusters. We also obtain a goose floating on the lake (named **NI2**) with 60 $\times$ 90 pixels from the Berkeley Segmentation Dataset [56]. The analysis of this natural gray image is given in *Section H of the supplementary file*. The clustering results of different methods based on these two images are given in Table II.



(a) Original data. (b) Result by BSC.

Fig. 8. The **FI5** dataset and the result of BSC on this database.

D. Experiment 4

In this experiment, the Olivetti Face Database [15], a widespread benchmark for machine learning, is applied to further evaluate the potential of the BSC in unsupervised image classification. The image data, called **FI5**, contains five people, each of whom has 10 face pictures with different shooting angles and expressions. The original data of these faces are shown in Fig. 8(a), where the faces of the same color belong to the same class, *i.e.* the same person. Here we set $Eps = 5$ and $MinPts = 3$ in DBSCAN. We take $f = 0.2$ in DPC and $r = 9$ is set in MS-type. $K = 8$, $\alpha = 3$, $\Delta = 10$ in BPEC and $\mathcal{K}_1 = 6$, $\alpha = 0.1$ are taken in BSC, respectively. The other parameters are default[3].

The clustering results of comparison methods are shown in *Section I of the supplementary file*[4]. We intuitively see that women's faces are identified accurately by these methods. However, for four other people, K-means, DBSCAN, MC, DPC, and MS-type misclassify them since some faces are not distinguished. For example, the faces of the fourth and fifth people are very similar, and these methods (MC, DPC, and MS-type) assign them to singleton clusters, which increases the risk of errors. Additionally, K-means and DBSCAN are very early clustering methods and provide bad performance on this dataset. MS-type is worse than MC and DPC since it is mainly suitable for low dimensional datasets while the dimension of **FI5** is high. We also see from Fig. 8(b),(f) that BPEC and BSC can produce the credal partitions to reduce errors. However, it is worth noting that BPEC assigns two faces marked by black of the fourth person to meta-cluster composed of the second and the fifth person. That is, BPEC argues that the two faces are difficult to distinguish between the second and the fifth person, which is unreasonable and results in worse performance than MC, DPC, and MS-type on this dataset. By contrast, BSC can effectively characterize the uncertainty and imprecision between singleton clusters with arbitrary shapes and sizes. Thus, it can recognize the first three people accurately. For two faces marked by cyan of the fifth people, BSC argues that they are hard to be distinguished between the fourth and the fifth people only using the current information. This is a prudent decision that can effectively reduce the error rates and fit what we reasonably expect. These imprecise images in meta-clusters can be eventually distinguished using some other techniques or with extra information sources. The clustering results of different methods are shown in Table III, and the results verify that BSC has potential in image classification. Additionally, some UCI real-world databases (available at http://archive.ics.uci.edu/ml/) are employed to evaluate the performance of BSC compared with other methods. These contents are included in *Section I of the supplementary file*.

[3]Here we take $\mathcal{K}_2 = 5$ since BSC needs to obtain different neighbors from various clusters of the precise objects to yield simulated cluster centers.

[4]The scale of **FI5** is small, and there is no result of U-SENC on this dataset since it mainly focuses on clustering extremely large-scale datasets.

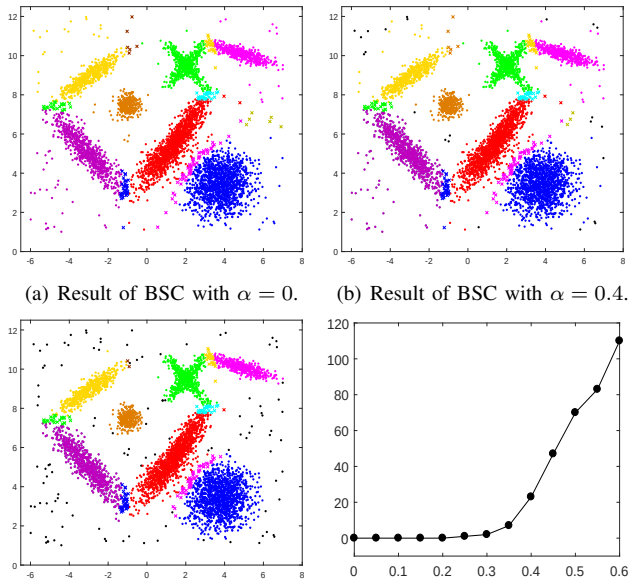
V. Discussion

The involved parameters. The 7-class dataset with arbitrary shapes and sizes is employed here to investigate the impact of parameters $(\alpha, \mathcal{K}_1, \mathcal{K}_2, \gamma)$ on the performance of BSC. The clustering results with three α values, *i.e.* $\alpha \in [0, 0.4, 0.6]$, are shown in Fig. 9(a)-(c), where the exact singleton clusters and proper meta-clusters marked by points and crosses with different colors, respectively. The black points represent the noise (outlier), and the number of noise with α from 0 to 0.6 is given by Fig. 9(d). The value of α corresponds to the x-coordinate, and the y-coordinate represents the number of noise. Interestingly, we can intuitively observe that BSC could assign all points to proper clusters without noise if we take a very small α, *e.g.* $\alpha = 0$, as shown in Fig. 9(a). As α increases, of course, those objects far from the clusters will be gradually assigned as noise, which is consistent with our intuitive perception. If we take a big α, *e.g.* $\alpha = 0.6$, some objects that are far from the clusters are regarded as noise, as shown in Fig. 9(c). We can also see that with the continuous increase of α, objects will increasingly be assigned as noise. Fig. 9 (d) reveals that the parameter α can effectively adjust the number of noise in BSC, and it depends on the number of noise that we can accept. The clustering results of BSC with various $\mathcal{K}_1$ and $\mathcal{K}_2$ are given in Figs. 10-11. We can see that the R_e, R_i, CRI values vary very little with different values

Table III: The results of different methods on the **FI5** dataset

Data	Indexes	K-means	DBSCAN	MC	DPC	MS-type	BPEC	BSC
FI5	$Re(\%)$	30.00	50.00	20.00	22.00	26.00	32.00	**14.00**
	$Ri(\%)$	/	/	/	/	/	0	**4.00**
	CRI	0.6986	0.3894	0.7278	0.5571	0.6251	0.7701	**0.9343**

of the parameters $\mathcal{K}_1$ and $\mathcal{K}_2$, which confirms its robustness to these two parameters. Particularly, we can see from Fig. 12 that the error rate R_e of BSC gradually decreases with the parameter γ changes from 0.5 to 2.5, while the imprecision rate R_i increases. This indicates that γ can reasonably adjust the degree of the imprecision and help to reduce the error rate. In applications, the parameter γ could correspond to a suitable compromise between the error rate and imprecision rate, and it depends on the imprecision rate that we can accept. In addition, the value of CRI is still stable even if γ is taken as different values, which also verifies the robustness and effectiveness of BSC.



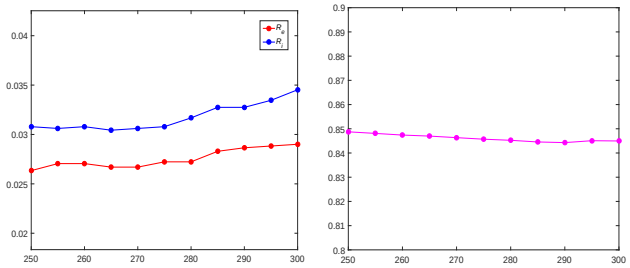
(a) Result of BSC with $\alpha = 0$. (b) Result of BSC with $\alpha = 0.4$.

(c) Result of BSC with $\alpha = 0.6$. (d) The noise with different α.

Fig. 9. The effect of α on the proposed BSC method.

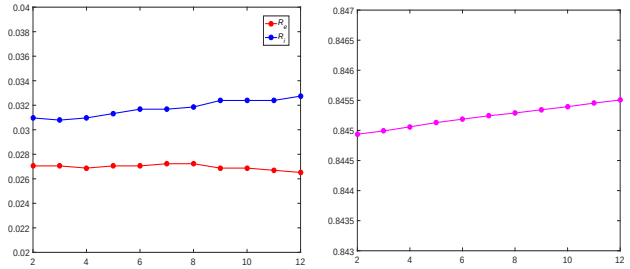
The complexity and other typical methods. We also discuss the complexity analysis of the proposed BSC method and investigate the execution time in seconds of BSC and other comparison methods on different UCI datasets. The detail is presented in *Section J of the supplementary file.* In addition, the differences and correlations between fuzzy-based methods, density-based methods, typical TBF-based clustering methods and the proposed BSC method are discussed in *Sections K and L of the supplementary file.*

The application of BSC in relational data. In this paper, the proposed BSC method mainly focuses on object data, and numerous experiments have verified its effectiveness. Different from object data, relational data (*e.g.* network data) only gives pairwise similarities or dissimilarities, and it has been widely applied to numerous applications, such as document classification [57], computational biology [58] and graph



(a) (b)

Fig. 10. (a) R_e and R_i values of BSC with different $\mathcal{K}_1$. (b) The CRI value of BSC with different $\mathcal{K}_1$.



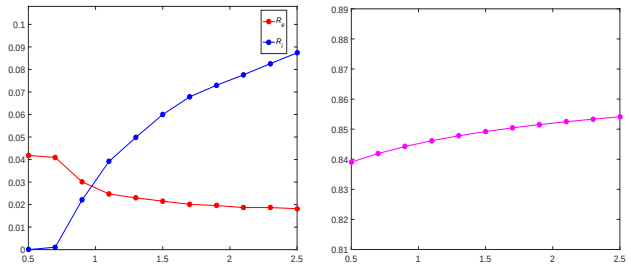
(a) (b)

Fig. 11. (a) R_e and R_i of BSC with different $\mathcal{K}_2$. (b) CRI of BSC with different $\mathcal{K}_2$.

Learning [59]. Some evidential clustering methods based on TBF have been developed for relational data. For example, a relational data-oriented method named EVCLUS is introduced in [41], extending TBF to clustering analysis for relational data. A median evidential c-means (MECM) clustering method is presented in [39], redefining the distance from the object to meta-clusters in ECM. In this way, MECM can apply to community detection. Inspired by these methods, BSC also has the potential for relational data. In future work, for example, we will try to apply BSC to graph clustering.

VI. Conclusion

In this paper, we propose a belief shift clustering (BSC) method to characterize the uncertainty and imprecision between the clusters with arbitrary shapes and sizes in the space, which can be regarded as the evidential version of mean shift or mode seeking under the TBF. In fact, all studies that include imprecise results, including our proposed BSC method, can be considered as prudent decision-making. Once an object is assigned to a meta-cluster, it indicates that this object cannot be precisely identified based on current knowledge. If this object is forced to be assigned to a singleton cluster, it may significantly increase the risk of error. In this case,



(a) (b)

Fig. 12. (a) R_e and R_i values of BSC with different γ. (b) The CRI value of BSC with different γ.

we need additional information to aid in further identification of these imprecise objects. Therefore, the results of the BSC can be considered as the first decision in applications. It aims to narrow down the number of objects identified and the potential solutions of imprecise objects. After obtaining imprecise objects, we can focus on acquiring new partial knowledge to identify them precisely. Although this would be costly, it is essential in some cases.

REFERENCES

[1] L. J. Ye, B. Liang, D. C. Yin and C. F. Yuan, "The K-Means-Type Algorithms Versus Imbalanced Data Distributions," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 728−745, Aug. 2012.

[2] Z. G. Su and T. Denœux, "BPEC: Belief-Peaks Evidential Clustering," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 111−123, 2019.

[3] J. C. Bezdek, J. M. Keller, "Streaming Data Analysis: Clustering or Classification?" *IEEE Trans. Syst., Man Cybern. Syst.*, vol. 51, no. 1, 2021.

[4] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651−666, 2010.

[5] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media, 2013.

[6] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98-110, Apr. 1993.

[7] N. R. Pal, K. Pal, J. M. Keller, J. C. Bezdek, "A possibilistic fuzzy *c*-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517-530, 2005.

[8] T. Lei, X. Jia, Y. Zhang, L. He and A. K. Nandi, "Significantly Fast and Robust Fuzzy C-Means Clustering Algorithm Based on Morphological Reconstruction and Membership Filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027-3041, 2018.

[9] H. Zhang, H. Li, N. Chen, S. Chen, J. Liu, "Novel fuzzy clustering algorithm with variable multi-pixel fitting spatial information for image segmentation," *Pattern Recognit.*, vol. 121, Jan. 2022. DOI: 10.1016/j.patcog.2021.108201.

[10] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, S. Qiu, C. T. Lin, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 1, pp. 40-52, 2021.

[11] G. Rayan, J. M. Keller, and M. Popescu, "TLPCM: Transfer Learning Possibilistic C-Means," *IEEE Trans. Fuzzy Syst.*, vol. 29,no. 4, pp. 940-952, 2020.

[12] Z. Deng, Y. Jiang, F. L. Chung, H. Ishibuchi, K. Choi, and S. Wang, "Transfer prototype-based fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1210-1232, 2015.

[13] R. N. Dave, "Clustering relational data containing noise and outliers," *Pattern Recognit Lett.*, vol. 12, pp. 657−664, 1991.

[14] F. Drabløs, "Symmetric distance measures for mass spectra," *Analytica chimica acta*, vol. 201, pp. 225−239, 1987.

[15] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science.*, vol. 344, no. 6191, pp. 1492−1496, 2014.

[16] Z. Bian, F. Chung, S. Wang, "Fuzzy Density Peaks Clustering," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 7, pp. 1725 - 1738, 2021.

[17] F. Camastra and A. Verri, "A novel kernel method for clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 801−805, 2005.

[18] H. Xiong, J. J. Wu and J. Chen, "K-means clustering versus validation measures: A data-distribution perspective," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 2, pp. 318-331, 2009.

[19] K. Fukunaga and L. D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Trans. Information Theory*, vol. 21, pp. 32−40, 1975.

[20] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790−799, 1995.

[21] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603−619, 2002.

[22] X. T. Yuan, B. G. Hu, R. He, "Agglomerative mean-shift clustering," *IEEE Trans. Knowledge and Data Eng.*, vol. 24, no. 2, pp. 209−219, 2010.

[23] M. Fashing and C. Tomasi, "Mean shift is a bound optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 471−474, 2005.

[24] S. Anand, S. Mittal, O. Tuzel and P. Meer, "Semi-Supervised Kernel Mean Shift Clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1201−1215, Jun. 2014.

[25] R. Yamasaki and T. Tanaka, "Properties of Mean Shift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.26, Apr. 2019.

[26] R. P. Duin, A. L. Fred, M. Loog and E. Pekalska, "Mode Seeking Clustering by kNN and Mean Shift Evaluated, in: Structural, Syntactic, and Statistical Pattern Recognition," *Springer.*, pp. 51-59, 2012.

[27] J. N. Myhre and K. Ø. Mikalsen, S. Løkse and R. Jenssen, "Robust clustering using a kNN mode seeking ensemble," *Pattern Recognit.*, vol. 76, pp. 491−505, 2018.

[28] M. Ester, H. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *in Proc. 2nd Int. Conf. Knowledge Discovery Data Mining*, pp. 226−231, 1996.

[29] Y. Chen, L. Zhou, S. Pei, Z. Yu, Y. Chen, X. Liu, J. Du, N. Xiong, "KNN-BLOCK DBSCAN: Fast clustering for large-scale data," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 6, Jun. 2021.

[30] A. J. Gallego, J. R. Rico-Juan, J. J. Valero-Mas, "Efficient k-nearest neighbor search based on clustering and adaptive k values," *Pattern Recognit.*, vol. 122, 2022. DOI: 10.1016/j.patcog.2021.108356.

[31] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395-416, 2007.

[32] D. Huang, C. Wang, J. Wu, J. Lai and C. Kwoh, "Ultra-Scalable Spectral Clustering and Ensemble Clustering," *IEEE Trans. Knowledge and Data Eng.*, vol. 32, no. 6, pp. 1212-1226, 2020.

[33] G. Shafer, "A Mathematical Theory of Evidence," *Princeton Univ. Press.*, 1976.

[34] P. Smets, "Analyzing the combination of conflicting belief functions," *Inf. Fusion*, vol. 8, no. 4, pp. 387−412, 2007.

[35] X. Deng, "Analyzing the monotonicity of belief interval based uncertainty measures in belief function theory," *Inter Jour Intell Syst.*, vol. 33, no. 9, pp. 1869−1879, 2018.

[36] Y. Yang and D. Han, "A new distance-based total uncertainty measure in the theory of belief functions," *Knowl.-Based Syst.*, vol. 94, pp. 114−123, 2016.

[37] M. H. Masson and T. Denœux, "ECM: an evidential version of the fuzzy *c*-means algorithm," *Pattern Recognit.*, vol. 41, no. 4, pp. 1364−1397, 2008.

[38] Z. G. Liu, Q. Pan, J. Dezert and G. Mercier, "Credal *c*-means clustering method based on belief functions," *Knowl.-Based Syst*, vol. 74, pp. 119−132, 2015.

[39] Z. Zhang, Z. Liu, A. Martin, Z. Liu and K. Zhou, "Dynamic evidential clustering algorithm," *Knowl.-Based Syst.*, 2021. DOI: 10.1016/j.knosys.2020.106643.

[40] Z. Zhang, Z. Liu, Z. Ma, Y. Zhang and H. Wang, "A new belief-based incomplete pattern unsupervised classification method," *IEEE Trans. Knowledge and Data Eng.*, 2021. DOI: 10.1109/TKDE.2021.3049511.

[41] T. Denœux and M. H. Masson, "EVCLUS: evidential clustering of proximity data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 95−109, Feb. 2004.

[42] C. Lian, S. Ruan, T. Denœux, H. Li and P. Vera, "Joint Tumor Segmentation in PET-CT Images Using Co-Clustering and Fusion Based on Belief Functions," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 755−766, Feb. 2019.

[43] Z. G. Liu, G. H. Qiu, G. Mercier, Q. Pan, "A transfer classification method for heterogeneous data based on evidence theory." *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 8, Aug. 2021.

[44] 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, 1995.

[45] Z. F. Ma, H. P. Tian, Z. C. Liu, and Z. W. Zhang, "A new incomplete pattern belief classification method with multiple estimations based on KNN," *Appl. Soft. Comput.*, vol. 90, 2020. DOI: 10.1016/j.asoc.2020.106175.

[46] Z. Zhang, H. Tian, L. Yan, A. Martin and K. Zhou, "Learning a credal classifier with optimized and adaptive multiestimation for missing data imputation," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 7, pp. 4092-4104, 2022.

[47] D. Han, J. Dezert, Y. Yang, "Belief interval-based distance measures in the theory of belief functions." *IEEE Trans. Syst., Man Cybern. Syst.*, vol. 48, no. 6, pp. 833-850, 2018.

[48] Z. G. Liu, X. Zhang, J. Niu, J. Dezert, "Combination of Classifiers With Different Frames of Discernment Based on Belief Functions," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 7, Jul. 2021.

[49] T. Denœux. "Decision-Making with Belief Functions: a Review," *Int. J. Approx. Reason.*, vol. 109, pp. 87−110, 2019.

[50] S. B. Hariz, Z. Elouedi, and K. Mellouli, "Selection initial modes for belief k-modes method," *Int. J. Appl. Sci., Eng. and Techn.*, vol. 4, no. 4, 2007.

[51] S. Haykin, "Neural Networks-A Comprehensive Foundation," *IEEE Press, New York*, 1994.

[52] T. Denœux, S. Li and S. Sriboonchitta, "Evaluating and comparing soft partitions: an approach based on Dempster-Shafer theory," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1231−1244, 2018.

[53] W. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846−850, 1971.

[54] P. Fränti and O. Virmajoki, Iterative shrinking method for clustering problems, *Pattern Recognit*. vol. 39, no.5, pp. 761−775, 2006.

[55] G. Argenziano, H. Soyer and V. D. Giorgi, "Dermoscopy: A Tutorial," *EDRA Medical Publishing and New Media*, 2002.

[56] D. Martin, C. Fowlkes, D. Tal and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *in: Proceedings of the 8th International Conference Computer Vision*, vol. 2, pp. 416−423, 2001.

[57] L. Hu, X. Pan, Z. Tan and X. Luo, "A Fast Fuzzy Clustering Algorithm for Complex Networks via a Generalized Momentum Method," *IEEE Trans. Fuzzy Syst.*, 2021. DOI: 10.1109/TFUZZ.2021.3117442.

[58] L. Hu, J. Zhang, X. Y. Pan, H. Yan, and Z. H. You, "HiSCF: leveraging higher-order structures for clustering analysis in biological networks," *Bioinformatics*, vol. 37, no. 4, pp. 542−550, 2021.

[59] R. Zhang, Y. Zhang, C. Lu and X. Li, "Unsupervised Graph Embedding via Adaptive Graph Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. DOI: 10.1109/TPAMI.2022.3202158.

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 the Doctoral degree in computer science from the Laboratory IRISA, University of Rennes 1, Lannion/Rennes, France, in 2022.

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

Zhun-ga Liu was born in Luoyang, China. He received the bachelor's and master's degrees from Northwestern Polytechnical University (NPU), Xi'an, China, in 2007 and 2010, respectively, and the Ph.D. degree from NPU and Telecom Bretagne, France, in 2013.

He is currently a Professor with the School of Automation, NPU. His research interest mainly focuses on pattern recognition, belief functions, and information fusion.

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.

Supplementary file

A. Some TBF-based clustering methods

A collection of mass function $m(.)$ for n objects is called credal partition [2]. It allows the object to be assigned to a singleton cluster thereby representing the exact information or meta-cluster composed of multiple singleton clusters thereby characterizing uncertainty and imprecision [37]-[42]. For example, Masson and Denœux first propose an evidential c-means (ECM) [37], regarded as the evidential version of the fuzzy c-means (FCM) [5] and noise clustering (NC) [13], which allows the object to be in any singleton clusters and any sets of some clusters, *i.e.* meta-cluster, with different masses of belief. An improved credal c-means (CCM) method is proposed in [38] to make ECM deal with the case when the centers of different clusters are very close. In previous work, we also propose a dynamic evidential clustering (DEC) in [39] to reduce the complexity of these typical TBF-based methods. Whereas these clustering methods based on the symmetric distance (FCM-like) may also not be suitable for the clusters with arbitrary shapes and sizes in the space. For example, if they are employed to deal with imbalanced data, the center of the meta-cluster tends to shift to the majority classes, which may result in unreasonable results [1]. Since the proposed BSC method is inspired in part by these methods such as ECM [37] and CCM [38], we also briefly review these two representative methods in *Sections B and C of this supplementary file*.

B. Brief review of evidential c-means (ECM)

Evidential c-means (ECM) is regarded as the evidential version of the fuzzy c-means (FCM) and noise clustering (NC) to characterize uncertainty and imprecision between different clusters, and it will be briefly introduced as follows.

Let us consider a dataset $\mathcal{X}$ including n objects with p attributes over the frame of discernment $\Omega = \{\omega_1, ..., \omega_c\}$. For the specific object $\mathbf{x}_i \in \mathcal{X}, i = 1, ..., n$, the mass of belief $m_{ij} \triangleq m_i(A_j)$ is associating the object $\mathbf{x}_i$ with an element A_j of the power-set 2^Ω. Particularly, $A_j \subseteq \Omega, A_j \neq \emptyset$, *i.e.* A_j can be any singleton cluster or meta-cluster included in 2^Ω. The cluster center $\bar{\mathbf{v}}_j$ associated to A_j can be computed by:

$$\bar{\mathbf{v}}_j = \frac{1}{|A_j|} \sum_{l=1}^{c} s_{lj} \mathbf{v}_l \tag{1}$$

subject to

$$s_{lj} = \begin{cases} 1, & \text{if } \{\omega_l\} \in A_j; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where $|A_j|$ denotes the cardinality of A_j and $\mathbf{v}_l$ represents the cluster center of $\{\omega_l\} \in \Omega$.

In ECM, the m_{ij} value depends on the distance d_{ij} between the object $\mathbf{x}_i$ and the cluster center $\bar{\mathbf{v}}_j$ of A_j, *i.e.* the higher distance d_{ij} leads to lower m_{ij}. ECM looks for the matrix M of the credal partition and the matrix V of cluster centers by minimizing the following objective function:

$$\mathcal{J}_{ECM}(M, V) = \sum_{i=1}^{n} \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} |A_j|^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\emptyset}^\beta \tag{3}$$

subject to

$$\sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1. \tag{4}$$

where the noise (outlier) threshold, denoted as δ, represents the distance between any object $\mathbf{x}_i$ $(i = 1, ..., n)$ and the noise cluster. $m_{i\emptyset}$ represents the mass of belief that the object assigned to the noise cluster and it can be adjusted by the threshold δ. A bigger threshold δ will lead to a lower mass of belief $m_{i\emptyset}$, and the object may be far away from the other objects if it is assigned to the noise cluster.

The object function $\mathcal{J}_{ECM}$ is then minimized by the Lagrange multipliers to provide the matrix M of the credal partition and the matrix V of cluster centers, defined by:

$$\begin{cases} m_{ij} = \frac{|A_j|^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} |A_k|^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}}, \text{ if } A_j \neq \emptyset; \\ m_{ij} = 1 - \sum_{A_j \neq \phi} m_{ij}, \text{ if } A_j = \emptyset. \end{cases} \tag{5}$$

where d_{ij} represents the distance between the $\mathbf{x}_i$ and the center of A_j. The exponent α controls the degree of penalization. β is a weighting exponent and generally set $\beta = 2$ as default.

The centers of the clusters are given by the rows of the matrix $V_{c\times p}$, given by:

$$V_{c\times p} = H_{c\times c}^{-1} B_{c\times p}; \tag{6}$$

subject to

$$B_{lq} = \sum_{i=1}^{n} \mathbf{x}_{iq} \sum_{\omega_l \in A_j} |A_j|^{\alpha-1} m_{ij}^\beta; \tag{7}$$

$$H_{lk} = \sum_{i=1}^{n} \sum_{\{\omega_l, \omega_k\} \subseteq A_j} |A_j|^{\alpha-2} m_{ij}^\beta. \tag{8}$$

where B_{lq} $(l \in [1, c], q \in [1, p])$ and H_{lk} $(l \in [1, c], k \in [1, c])$ represent the elements in the matrices $B_{c\times p}$ and $H_{c\times c}$, respectively.

ECM repeatedly updates the matrix M of the credal partition and the matrix V of cluster centers until the deviation between the two consecutive matrix V cluster centers after t iterations is less than the value of the threshold ε, *i.e.* $\|V_t - V_{t-1}\| < \varepsilon$.

C. Brief review of credal c-means clustering (CCM)

Compared with ECM, the mass of belief that the object $\mathbf{x}_i$ assigned to the meta-cluster in CCM depends not only on the distance from $\mathbf{x}_i$ to the meta-cluster center but also the distance between $\mathbf{x}_i$ and the singleton clusters included in the meta-cluster. The CCM can avoid the unreasonable result provided by ECM in clustering the datasets with special distributions. Additionally, ECM considers all the meta-clusters in the power-set 2^{Ω}. In such a case, it will bring a high computational complexity if the dataset contains a large number of clusters. Moreover, the CCM also sets a threshold $t_c \in [2, 2^{\Omega}]$ to eliminate some meta-clusters with big cardinality to reduce the computational complexity, especially in the datasets with abundant clusters. In this case, the set of the selected available clusters S^{Ω} is given by $S^{\Omega} = \{A_j, |A_j| < t_c\}$ in CCM.

Based on the above principles, the objective function $\mathcal{J}_{CCM}$ of CCM is defined by:

$$\mathcal{J}_{CCM}(M,V) = \sum_{i=1}^{n} \sum_{j/A_j \in S^{\Omega}} m_{ij}^{\beta} D_{ij}^2, \tag{9}$$

subject to

$$\sum_{j/A_j \in S^{\Omega}} m_{ij} = 1. \tag{10}$$

and

$$D_{ij}^2 = \begin{cases} \delta^2, & \text{if } |A_j| = \emptyset; \\ d_{ij}^2, & \text{if } |A_j| = 1; \\ \frac{\sum\limits_{A_l \in A_j} d_{il}^2 + \gamma d_{ij}^2}{|A_j| + \gamma}, & \text{if } |A_j| > 1. \end{cases} \tag{11}$$

where d_{ij} is the Euclidean distance between $\mathbf{x}_i$ and the center of the cluster A_j. d_{il} represents the distance from $\mathbf{x}_i$ to the centers of singleton clusters in the meta-cluster A_j such that $|A_j| > 1$. γ is the weighting factor of the distance between the object and the meta-cluster center, and it is used to control the imprecision rate. The bigger the value of γ is, the more objects will be assigned to the meta-clusters, and it is generally taken $\gamma \in [0.5, 3]$. The weighting exponent $\beta = 2$ is the default.

The Lagrange multipliers method is used to minimize the function $\mathcal{J}_{CCM}(M,V)$ to obtain the matrix M of the credal partition and the matrix V of cluster centers, defined by:

$$m_{ij} = \frac{D_{ij}^{-2/(\beta-1)}}{\sum\limits_{k|A_k \in S^{\Omega}} D_{ik}^{-2/(\beta-1)}} \tag{12}$$

The centers of the cluster are given by the rows of the matrix $V_{c\times p}$, given by:

$$V_{c\times p} = H_{c\times c}^{-1} B_{c\times n} \mathcal{X}_{n\times p}; \tag{13}$$

subject to

$$\begin{aligned} B_{li} &= m_{il}^{\beta} + \sum_{A_l \in A_j} m_{ij}^{\beta} \tfrac{1+\gamma}{|A_j|+\gamma}; \\ H_{ll} &= \sum_{i=1}^{n} m_{il}^{\beta} + \sum_{i=1}^{n} \sum_{A_l \in A_j} m_{ij}^{\beta} \tfrac{1+\frac{\gamma}{|A_j|^2}}{|A_j|+\gamma}; \\ H_{lq} &= \sum_{i=1}^{n} \sum_{\{A_l, A_q\} \in A_k} m_{ij}^{\beta} \tfrac{\gamma}{|A_k|^2(|A_k|+\gamma)}, \quad l \neq q. \end{aligned} \tag{14}$$

where B_{li} ($l \in [1,c]$, $i \in [1,n]$) and H_{lq} ($l \in [1,c]$, $q \in [1,c]$) represent the elements in the matrix $B_{c\times n}$ and $H_{c\times c}$, respectively.

CCM repeatedly updates the matrix M of the credal partition and the matrix V of cluster centers, and the termination condition is the same as that of ECM.

D. Basics of Mean shift

Mean shift (MS) is first introduced by Fukunaga and Hostetler [19]-[20]. They designed a simple nonparametric iterative procedure that shifts each object to the average of the objects included in its neighborhood. We briefly review the generalized mean shift procedure as follow.

Let us consider that S is a p-dimensional Euclidean space, and the dataset, named $\mathcal{X}$ with $\mathcal{X} \subset S$, is a finite set. The object $\mathbf{s}_i \in S$ as the initial cluster center (mode) and the new cluster center $\mathbf{s}_{i+1}$ which $\mathbf{s}_i$ shifts to is given by:

$$\mathbf{s}_{i+1} = \frac{\sum\limits_{\mathbf{x}_j \in \mathcal{X}} K(\mathbf{x}_j - \mathbf{s}_i)\mathbf{x}_j}{\sum\limits_{\mathbf{x}_j \in \mathcal{X}} K(\mathbf{x}_j - \mathbf{s}_i)} \tag{15}$$

with

$$K(\mathbf{x}_j - \mathbf{s}_i) = \begin{cases} 1, & \text{if } \|\mathbf{x}_j - \mathbf{s}_i\| \leq h \\ 0, & \text{if } \|\mathbf{x}_j - \mathbf{s}_i\| > h \end{cases} \tag{16}$$

where h is called the bandwidth and $||.||$ represents the Euclidean distance. $K(\mathbf{x}_j - \mathbf{s}_i)$ is the *unit flat kernel* here and it can also be other kernels (*e.g.* Gaussian kernel [20]). The cluster center $\mathbf{s}_i$ shifts to the new cluster center $\mathbf{s}_{i+1}$, denoted as $\mathbf{s}_i \leftarrow \mathbf{s}_{i+1}$. According to the mean shift vector, named $\mathbf{m}(\mathbf{s}_i)$, it can be concluded by $\mathbf{m}(\mathbf{s}_i) = \mathbf{s}_{i+1} - \mathbf{s}_i$. The MS method repeatedly updates the cluster center $\mathbf{s}_i$ using Eqs. (15)−(16) until the deviation between the estimations of two consecutive cluster centers is less than the value of the threshold ε, *i.e.* $\|\mathbf{s}_{i+1} - \mathbf{s}_i\| < \varepsilon$. In the process of iterations, each object will be searched by other ones that coverage to one or more clusters. The object is searched by one or multiple clusters, and it will be assigned to the cluster that searches it the most times. The flowchart of mean shift (MS), as shown in Fig. 1, illustrates its basic principle intuitively.

E. Some knowledge of the proposed BSC method

The flowchart of preliminary adaptive credal partition by belief shift is shown in Fig. 2.

F. Parameters in experiments

All parameters are defaults except the ones we adjust to make the experiments intuitive and concise. Specifically, DBSCAN mainly contains two parameters Eps and $MinPts$ that are used to assign each object to noise or a singleton cluster. Inspired by [28], the K-distance graph method is employed to obtain the Eps value, and the $MinPts$ value such that $MinPts \geqslant 3$. In DPC, the parameter f is used to determine the cutoff distance and recommended $f \in [0.01, 0.02]$. MS-type has the parameter r that represents the bandwidth. The higher (lower) the difference between attributes, the bigger

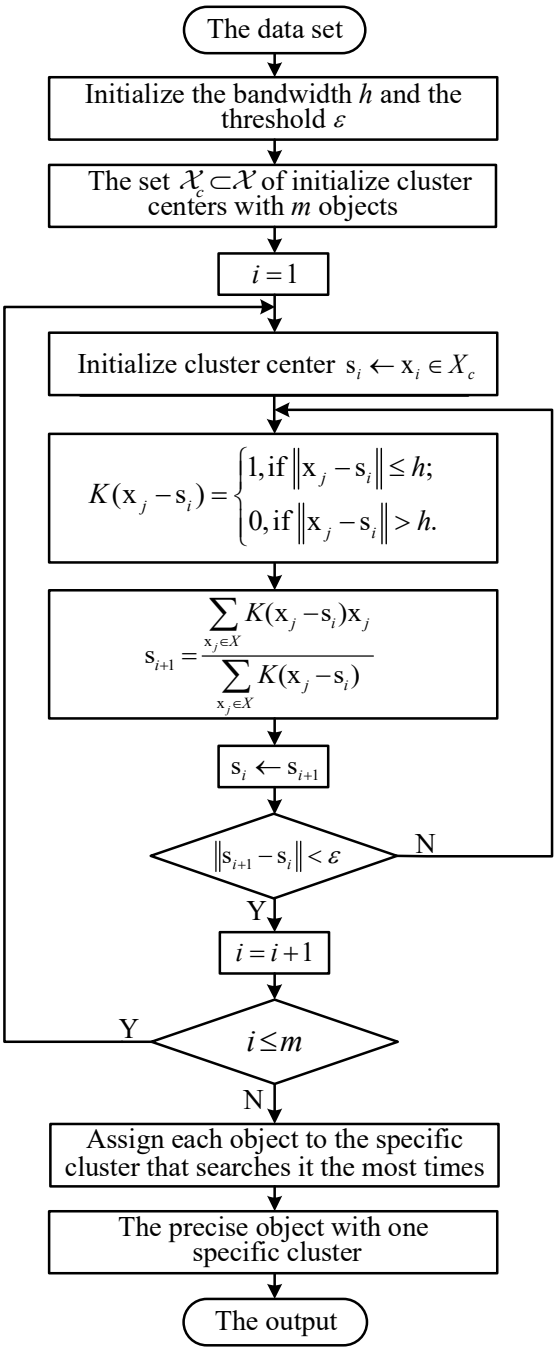


Fig. 1. Illustration of the principle of mean shift procedure.

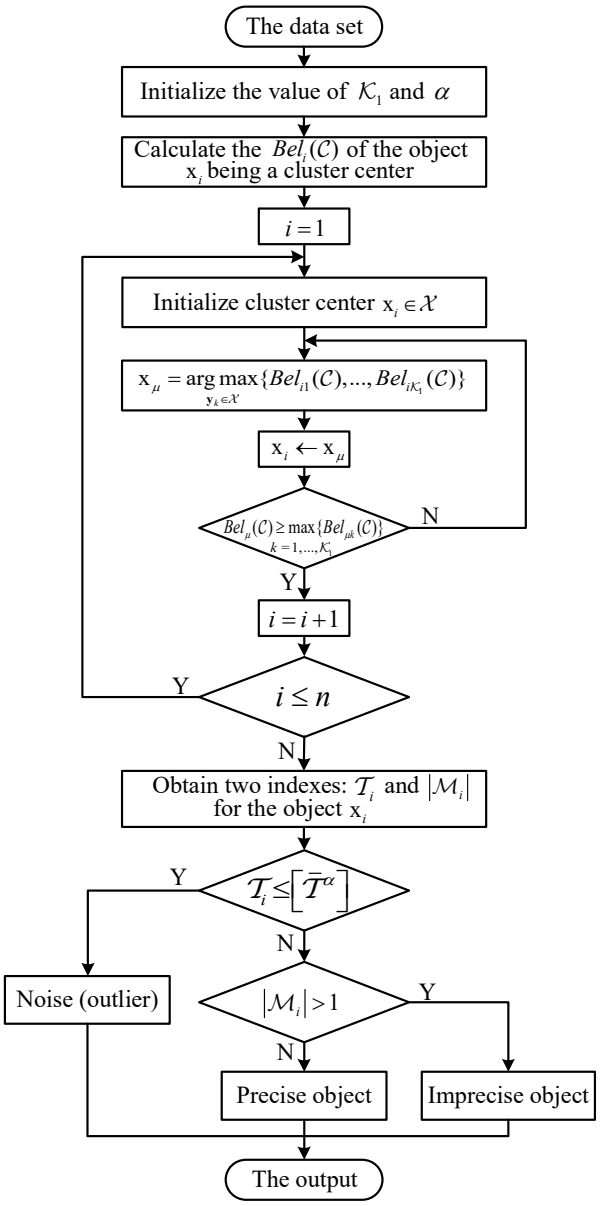
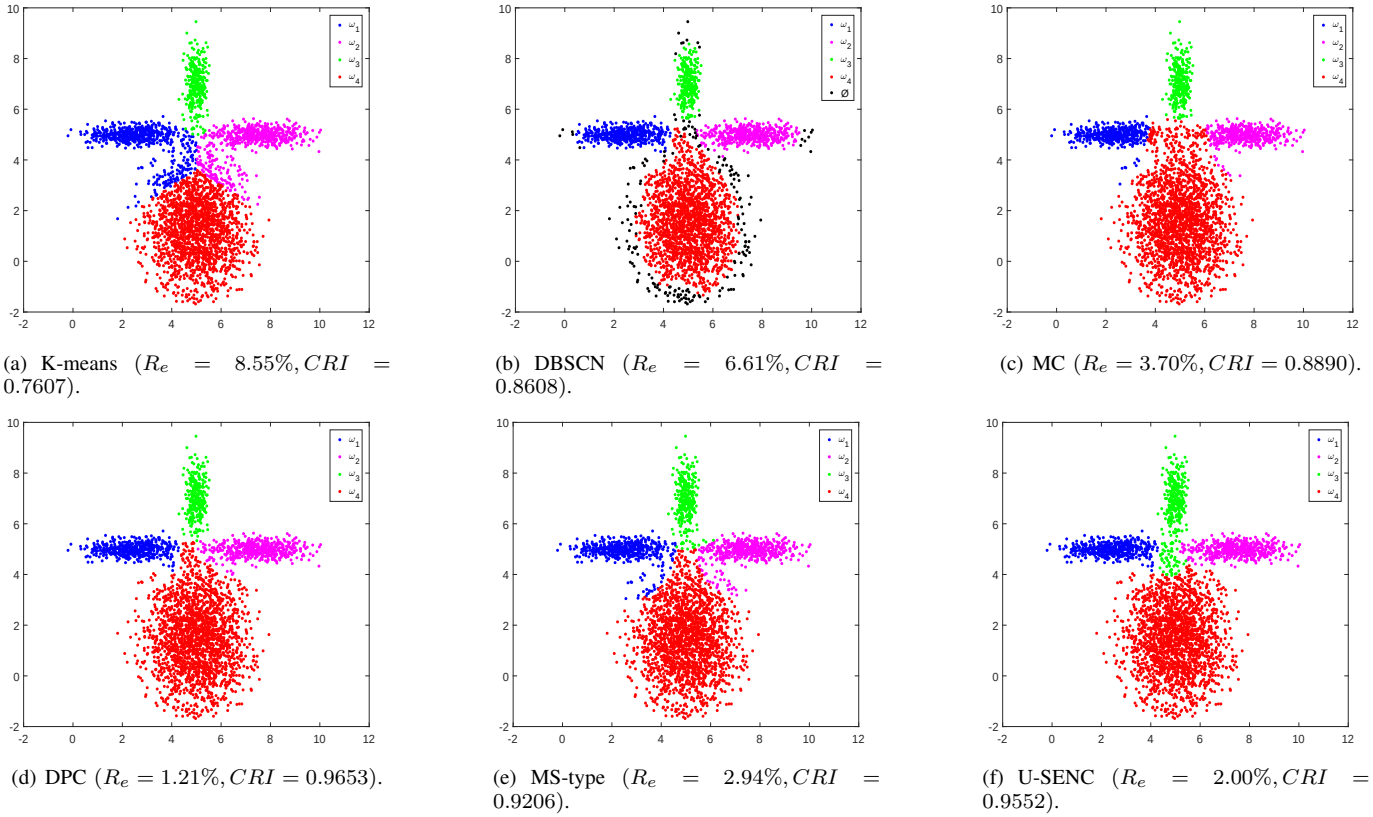


Fig. 2. Illustration of the principle of preliminary adaptive credal partition.

(smaller) r is given. In BPEC, there are three mainly parameters K, α, Δ. K represents the number of neighbors that used to conclude the degree of belief and a large value of K is preferable for most datasets. The parameter α, similar to that in ECM [37], refers to the weighting exponent for cardinality, such that $\alpha \geqslant 0$. Δ is used to assign noise and is equal to a constant smaller than the minimal delta associated to noise in the decision graph.

G. **Experiment 2**

Fig. 7(a)-(f) shows the clustering results of different methods on the **SD4** dataset. From Fig. 7(a), we can see that K-means has poor performance because it assigns a part of objects belonging to the majority class to the minority classes, which makes clusters have relatively uniform sizes. In Fig. 7(b), DBSCAN assigns some objects that originally belong to singleton clusters into noise cluster, which brings error partition. The results of MC, DPC and MS-type are given in Fig. 7(c)-(e). Although they can effectively cluster such a dataset, they cannot assign the objects distributed in the overlapping areas. From Fig. 7(f), we can see that U-SENC mainly focuses on the scalability and robustness of spectral clustering for extremely large-scale datasets with limited resources. It also provides poor performance in assigning the objects in overlapping areas of different clusters.



(a) K-means ($R_e = 8.55\%, CRI = 0.7607$). (b) DBSCN ($R_e = 6.61\%, CRI = 0.8608$). (c) MC ($R_e = 3.70\%, CRI = 0.8890$).

(d) DPC ($R_e = 1.21\%, CRI = 0.9653$). (e) MS-type ($R_e = 2.94\%, CRI = 0.9206$). (f) U-SENC ($R_e = 2.00\%, CRI = 0.9552$).

Fig. 7. The clustering results of different methods on the **SD4** dataset.

H. Experiment 3

Fig. 8 shows the segmentation results of the comparison methods on the (**DI2**) dataset. Here we also employ a natural gray image of 60×90 pixels, named **NI2**, a goose floating on the lake, to estimate the performance of different methods, as shown in Fig. 8+ (a). This image consists of the goose and the background, and Fig. 8+ (c) shows the ground truth. We give the distribution of the gray value of each pixel in this image, as shown in Fig. 8+ (b). The x-coordinate represents the number of different pixel and y-coordinate is the gray value from 0 to 255. The pixels in goose and background are marked by blue and red points, respectively. Here we set $Eps = 170$ and $MinPts = 0.1$ in DBSCAN. We take $f = 0.02$ in DPC and the bandwidth $r = 120$ in MS-type. $K = 1500$, $\alpha = 5$, $\Delta = 500$ and $\mathcal{K}_1 = 1700$ are set in BPEC and BSC, respectively.

Fig. 8+ reports the clustering results of different methods based on the **NI2** dataset. From Fig. 8+ (a)-(b), we can observe that some pixels are indistinguishable for the goose and background depending only on the gray values. For example, the reflection in the water is very similar to the goose, and it is challenging to assign these pixels. Some of these pixels are forced to be wrongly assigned to the goose cluster by K-means, DBSCAN, MC, DPC, MS-type and U-SENC, whereas BPEC and BSC can effectively reduce the error by assigning these pixels to the meta-cluster. Additionally, the clustering results of different methods are given in Table 2 in the manuscript, where BSC yields lower error rates, acceptable imprecision rates, and higher CRI. It indicates that the performance of BSC is superior to other methods.

I. Experiment 4

Fig. 9 reports the clustering results of the comparison methods on the Olivetti Face Database (**FI5**).

In this experiment, we evaluate the performances of BSC compared with other methods based on the UCI databases (available at http://archive.ics.uci.edu/ml/). Table 4 reports the basic information of these datasets including the number of clusters (#Clus.), attributes (#Attr.), instances (#Inst.) and the number of objects in different classes.

The main parameters of DBSCAN, DPC, MS-type, and BPEC are given in Table 5 and the other parameters in this experiment are default. Here the cluster label of these datasets doesn't contain noise and we take proper Δ in BPEC and $\alpha = 0$ in BSC, respectively. That is, the clustering results of different methods do not include noise. The clustering results of different methods are shown in Table 6. From the experimental results, we can see that the error of BSC is obviously lower than that of other methods in most cases, and the imprecision rates are within an acceptable range, which can truly reflect that BSC is superior to other methods.

J. Complexity analysis

Let us consider a dataset $\mathcal{X}$ including n objects over the frame of discernment $\Omega = \{\omega_1, ..., \omega_c\}$. The computational complexity of BSC mainly comes from the calculation of

Table 4: Basic information of the UCI datasets

Datasets	#Clus.	#Attr.	#Inst.	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_4\}$	$\{\omega_5\}$	$\{\omega_6\}$
Appendicitis(Ap)	2	7	106	21	85	/	/	/	/
Biodeg(Bi)	2	41	1055	356	699	/	/	/	/
Spambase(Sp)	2	56	4597	1813	2788	/	/	/	/
Facebook(Fa)	2	9	6622	2334	4288	/	/	/	/
Hill valley (Hv)	2	100	606	311	295	/	/	/	/
Abalone(Ab)	3	7	600	233	239	128		/	/
Seeds(Se)	3	7	210	70	70	70	/	/	/
Contraceptive(Co)	3	8	1473	629	333	511	/	/	/
Sensor (Sen)	4	24	5456	826	2097	2205	328	/	/
Urban land cover (Ulc)	5	147	464	116	61	106	59	122	/
Red wine quality (Rwq)	6	11	1599	10	53	681	638	199	18

Table 5: Selection of the parameters in different methods

Methods	DBSCAN		DPC	MS-type	BPEC			BSC
Indexes	Eps	MinPts	f	r	K	α	Δ	$\mathcal{K}_1$
Ap	1	13	0.5	0.55	50	2	30	10
Bi	2.5	12	0.02	70	100	2	200	400
Sp	0.5	50	0.02	12000	200	2	5000	2000
Fa	0.2	110	0.02	0.6	100	2	500	400
Hv	2	20	0.02	1.2	50	5	50	33
Ab	1	60	0.02	3	30	6	30	60
Se	1	20	0.02	2	20	3	30	20
Co	0.8	12	0.02	7	30	2	30	200
Sen	0.5	30	0.02	2	100	2	600	350
Ulc	2	9	0.02	2.4	30	4	60	15
Rwq	2	15	0.02	0.23	120	2	500	40

Table 6: Clustering results of different methods with the UCI datasets (Re/Ri in %)

Datasets	Indexes	K-means	DBSCAN	MC	DPC	MS-type	BPEC	U-SENC	BSC
Ap	Re	18.87	19.81	24.53	13.21	20.75	19.81	31.13	**12.26**
	Ri	/	/	/	/	/	2.83	/	**2.83**
	CRI	0.3348	0.0000	0.2410	0.4194	-0.0139	0.7127	0.0481	**0.7795**
Bi	Re	41.14	34.31	37.63	33.84	34.60	41.99	38.20	**32.13**
	Ri	/	/	/	/	/	2.09	/	**13.65**
	CRI	0.0037	-0.0055	-0.0268	-0.0009	-0.0081	0.6121	0.0544	**0.7138**
Sp	Re	36.41	40.90	37.11	35.81	39.37	33.41	39.06	**28.52**
	Ri	/	/	/	/	/	4.74	/	**7.50**
	CRI	0.0394	-0.0096	0.0077	0.0450	0.0005	0.5927	0.0039	**0.6426**
Fa	Re	33.16	34.22	34.32	34.76	35.23	35.91	33.80	**33.13**
	Ri	/	/	/	/	/	0.50	/	**3.47**
	CRI	0.0364	0.0202	0.0158	0.0450	0.0003	0.5492	0.0259	**0.5700**
Hv	Re	48.51	48.68	48.02	48.68	49.50	47.52	47.69	**47.03**
	Ri	/	/	/	/	/	1.98	/	**3.80**
	CRI	0.0001	0.0000	0.0006	0.0009	0.0007	0.5191	0.0006	**0.5261**
Ab	Re	50.67	60.67	55.00	54.50	60.17	57.17	58.17	**52.17**
	Ri	/	/	/	/	/	5.00	/	**0**
	CRI	0.0741	-0.0058	0.0455	0.0985	-0.0008	**0.5896**	0.0152	0.5570
Se	Re	10.95	33.33	18.10	11.43	13.33	8.57	12.38	**7.62**
	Ri	/	/	/	/	/	6.19	/	**7.62**
	CRI	0.7103	0.2838	0.5702	0.7027	0.6592	0.8242	0.6825	**0.9027**
Co	Re	59.88	56.89	61.10	59.06	60.56	57.98	61.78	**56.96**
	Ri	/	/	/	/	/	7.26	/	**6.65**
	CRI	0.0257	0.0037	0.0236	0.0099	0.0186	**0.6861**	0.0185	0.6244
Sen	Re	60.03	59.97	59.82	62.02	59.59	55.96	55.96	**54.22**
	Ri	/	/	/	/	/	0.00	/	**16.42**
	CRI	0.0562	-0.0015	0.0563	0.0232	0.0000	0.5936	0.0691	**0.8441**
Ulc	Re	69.61	73.71	65.09	64.44	52.16	31.25	67.03	**24.35**
	Ri	/	/	/	/	/	0.00	/	**8.62**
	CRI	0.0495	0.0000	0.0795	0.0277	0.3873	0.7485	0.0702	**0.7969**
Rwq	Re	70.98	59.16	70.36	61.85	63.29	65.35	76.17	**58.66**
	Ri	/	/	/	/	/	1.63	/	**15.95**
	CRI	-0.0020	0.0173	0.0166	0.0342	0.0779	0.6853	0.0064	**0.7862**

Table 7: Execution time of different methods (In seconds)

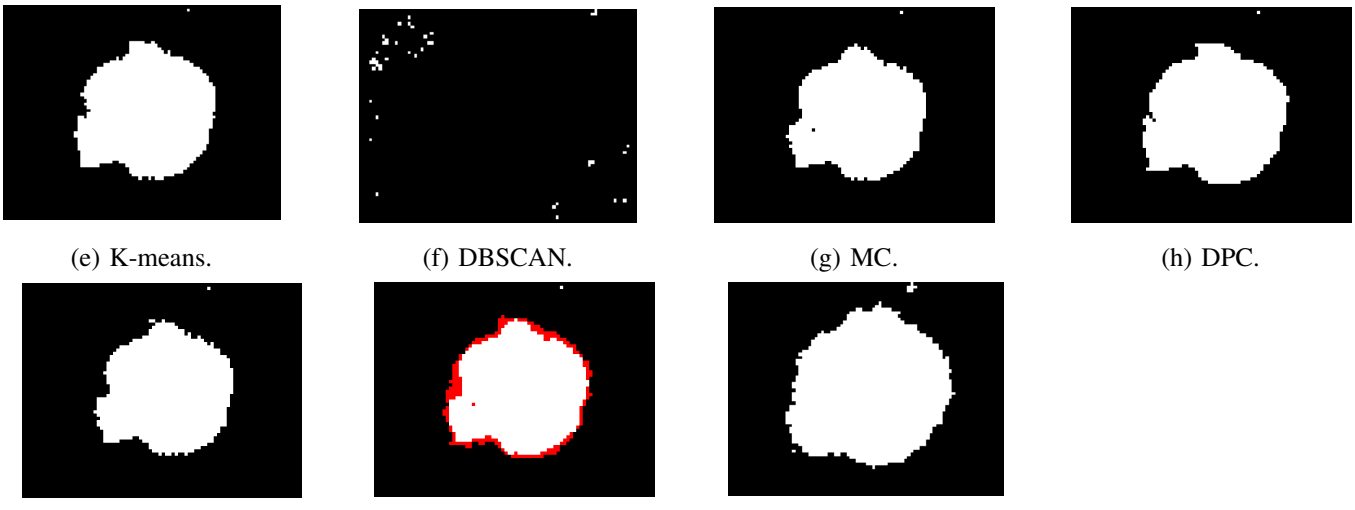
Datasets	K-means	DBSCAN	MC	DPC	MS-type	BPEC	U-SENC	BSC
Ap	0.1356	0.0511	0.0832	0.8196	0.0580	0.4885	0.9462	0.1662
Bi	0.1682	0.1482	0.2320	3.4225	0.1383	0.4944	5.0269	0.4538
Sp	0.3324	3.6204	1.2315	4.2871	0.6112	2.0783	13.8553	6.5877
Fa	0.1957	2.4191	1.3657	11.7816	6.5137	2.0067	13.6544	13.2586
Hv	0.2240	0.2150	0.3156	1.2701	0.2515	2.5780	5.5430	4.8632
Ab	0.1442	0.1687	0.1489	0.8970	0.1291	1.2620	2.1973	0.1606
Se	0.1346	0.0334	0.1529	0.8311	0.0557	0.6525	1.2732	0.1354
Co	0.1465	0.1289	0.3182	1.0405	0.3574	1.9261	5.3182	0.4992
Sen	0.3116	2.7066	2.5923	6.6450	2.9639	4.4734	11.9310	8.1370
Ulc	0.2248	0.1394	0.5450	1.3616	0.7835	2.3256	2.4134	2.2083
Req	0.1932	0.2118	0.9816	8.0237	2.0427	0.5940	6.5836	1.1436

Euclidean distances between objects to find neighbors. In the first step, each object needs to search the KNNs from $\mathcal{X}$ to compute its degree of belief. Thus, the computational complexity of this process is $\mathcal{O}(n^2)$. In the second step, there are n_1 imprecise objects are credal redistributed by the evidential clustering rule. For the query object $\mathbf{x}_i$, it needs to search different KNNs from clusters included in its dynamic sub-framework $\mathcal{M}_i$ ($|\mathcal{M}_i| > 1$) thereby simulating the singleton cluster centers. One can consider that there are n_i objects in these singleton clusters, and the computational complexity of this process is $\mathcal{O}(\sum_{i=1}^{n_1} n_i)$. Therefore, the total computational complexity of BSC is $\mathcal{O}(n^2 + \sum_{i=1}^{n_1} n_i)$.

In addition, to compare the computational complexity with other methods, we investigate the execution time in seconds of BSC and other comparison methods on different UCI datasets. The execution time in seconds of BSC and other comparison methods on the different datasets are shown in Table 7. We can see that the execution time of BSC is higher than some methods since it needs to calculate a large number of distances between objects to obtain the neighbors $\mathcal{K}_1$ and $\mathcal{K}_2$ in the two steps. However, the BPEC and U-SENC methods take a long time on training and optimizing models, so the BSC method runs considerably faster than them in most cases. In applications, the proposed BSC method is more suitable for cases where high accuracy is required, whereas efficient computation is not a vital requirement.

K. The BSC method v.s. Fuzzy and density-based methods

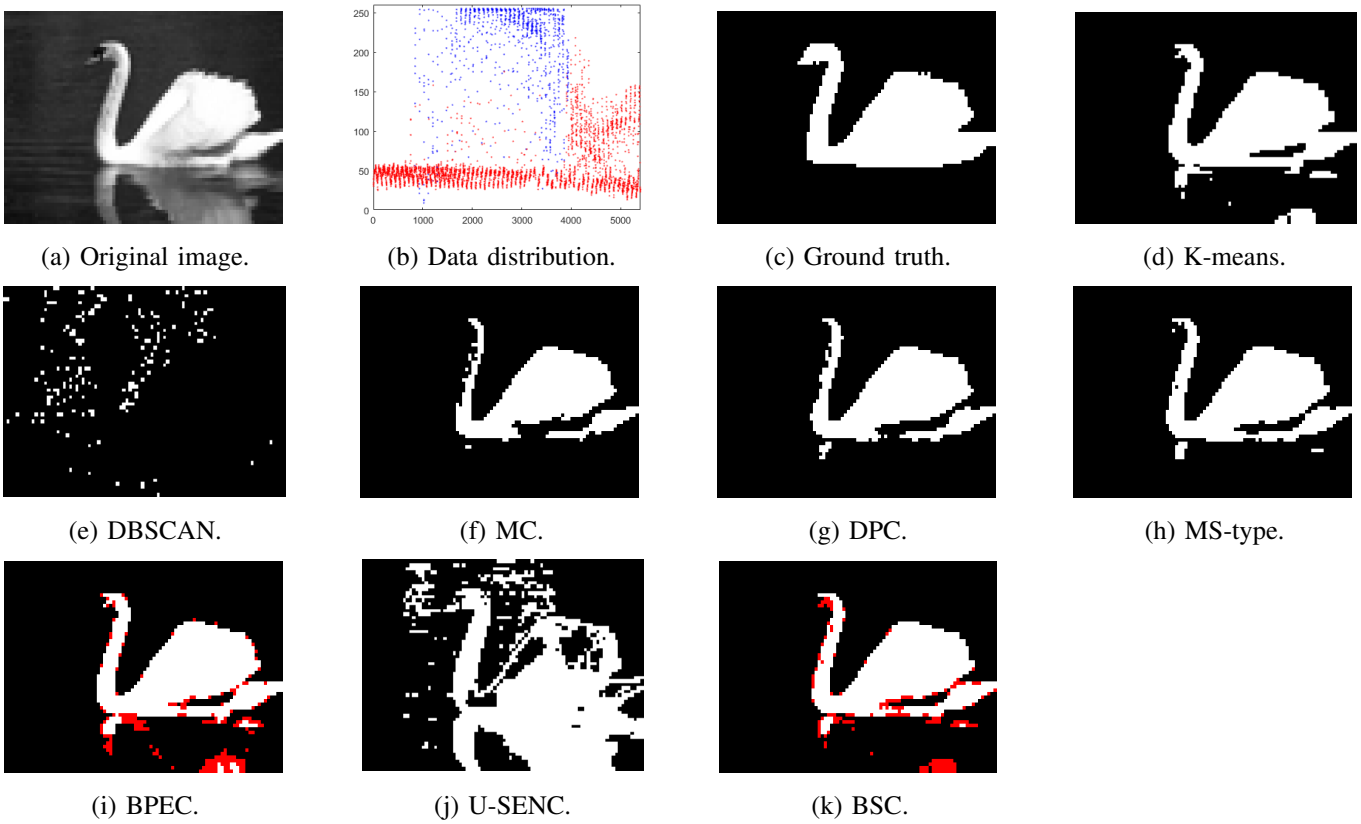
We compare the differences and correlations between fuzzy-based methods, density-based methods, and the proposed BSC method here. 1) It is well known that fuzzy-based methods, such as FCM, can provide a membership matrix that contains all objects' behavior to all clusters. This advantage is hard to reach by other types of clustering methods, which also provides the possibility for subsequent analysis of objects. For example, the obtained fuzzy data about objects can be further analyzed in the scene including clustering and data fusion. However, similar to K-means, most fuzzy-based methods are considered partition-based methods. Clusters are groups of data characterized by a small distance to the cluster center. An objective function, typically the sum of the distance to a set of putative cluster centers, is optimized until the best candidates are found [15]. In this case, these methods cannot detect nonspherical clusters because objects are always assigned to the nearest centers. 2) Unlike fuzzy-based methods, it is argued that density-based methods, such as DBSCAN and mean shift, can easily detect clusters with arbitrary shape and



(e) K-means. (f) DBSCAN. (g) MC. (h) DPC.

(i) MS-type. (j) BPEC. (k) U-SENC.

Fig. 8. The clustering results of different methods on the **DI2** dataset.



(a) Original image. (b) Data distribution. (c) Ground truth. (d) K-means.

(e) DBSCAN. (f) MC. (g) DPC. (h) MS-type.

(i) BPEC. (j) U-SENC. (k) BSC.

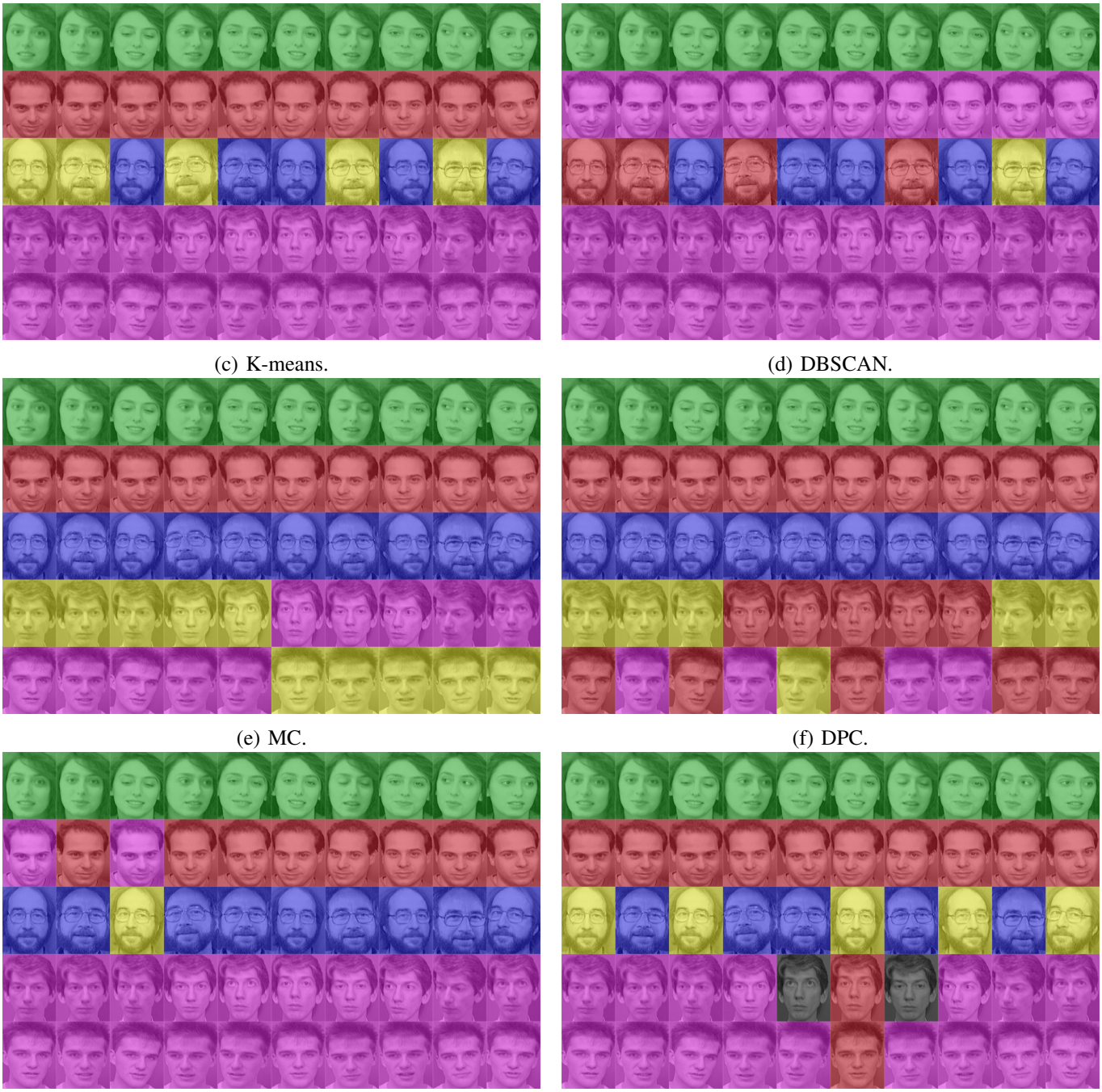
Fig. 8+. The clustering results of different methods on the **NI2** dataset.

size. However, once objects are assigned to different specific clusters, we cannot obtain uncertain and imprecise information about the objects (especially those in the overlapping or intermediate areas of different clusters) as we can with fuzzy-based methods. 3) The proposed BSC method in this paper combines the advantages of fuzzy-based and density-based methods. It aims to characterize the uncertainty and imprecision between clusters with arbitrary shapes and sizes. Specifically, we first assign each object as the noise, precise object, or imprecise object based on the notion of "belief shift". This step can be considered an inheritance of the advantages of density-based methods because arbitrary clusters can be easily detected by doing so. Then, we employ the evidential clustering rule to reassign imprecise objects to singleton clusters or associated meta-cluster. In this step, each imprecise object will get the behavior to the associated singleton or meta-clusters, *i.e.* mass function $m(\cdot)$. In fact, evidential clustering is considered the evidential version of FCM and NC, and $m(\cdot)$ is viewed as the improvement of membership function $u(\cdot)$. Therefore, the proposed BSC method also combines the advantage of fuzzy-based methods. The difference is that BSC provides uncertainty and imprecision for all imprecise objects, while fuzzy-based methods only analyze the uncertainty for all objects. However, the proposed BSC also has a few potential issues in applications. For example, the BSC needs many distance (similarity) calculations in each step to be time-consuming. Thus, it is suitable for prudent decision-making, whereas efficiency is unnecessary. Therefore, in the future, we will combine some more reasonable KNNs procedures [29], [30] with improving execution efficiency.

L. The BSC method v.s. Typical TBF-based methods

Here we compare the differences and correlations between BSC and other typical TBF-based methods. These methods have the advantages of characterizing uncertainty and imprecision between clusters. However, similar to fuzzy-based methods, most typical TBF-based methods are only considered partition-based methods. In other words, they cannot detect nonspherical clusters because objects are always assigned to the nearest centers. By contrast, the proposed BSC is considered the evidential version of mean shift under the TBF, so it has the advantage of density-based methods and can detect clusters with arbitrary shapes and sizes. Moreover, the BSC reassigns imprecise objects by a specific dynamic sub-framework with corresponding simulated centers rather than the fixed cluster centers used by most typical TBF-based methods, effectively avoiding the "uniform effect".

Based on the above analysis, the 7-class dataset with arbitrary shapes and sizes as shown in Fig. 10(a) is employed to investigate the performance of BSC compared to other typical TBF-based methods including ECM [37], CCM [38], EVCLUS [41] and BPEC [2]. In Fig. 10(a), various colors mark the data points included in different clusters, and noisy points are marked by black. Here, we set $\alpha = 2$ and $\gamma = 1$ in ECM and CCM, respectively. We choose $K = 500$, $\alpha = 2$, $\Delta = 5$ in BPEC. $K = 250$ is set in BSC. The other parameters



(c) K-means. (d) DBSCAN.

(e) MC. (f) DPC.

(g) MS-type. (h) BPEC.

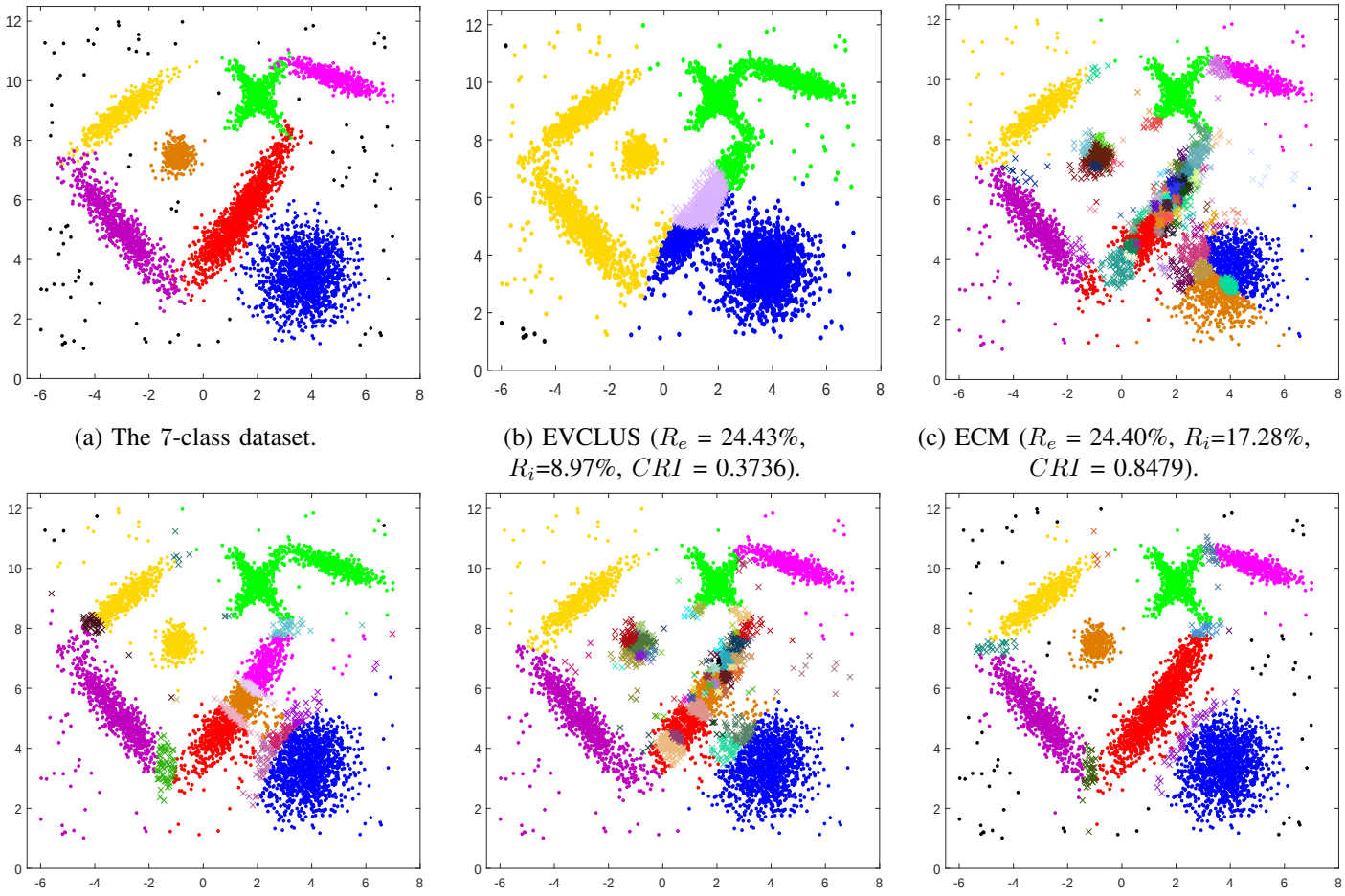
Fig. 9. The clustering results of different methods on the **FI5** dataset.

are defaults. The clustering results are reported in Fig. 10(b)-(f), where the singleton clusters and meta-clusters are marked by points and cross with different colors, respectively. We can see that these typical TBF-based methods cannot detect nonspherical clusters and noisy points. Moreover, they fail to assign the imprecise objects to reasonable meta-clusters since the centers of meta-clusters are incorrect. By contrast, the proposed BSC can effectively distinguish noisy points and different clusters with arbitrary shapes and sizes. It can also reasonably characterize the uncertainty and imprecision between arbitrary clusters, exhibiting better performance than other typical TBF-based methods.

M. Definition and proofs

Here we provide some definitions and proofs related to Eqs. (7)-(9). We first give some basic concepts as follows.

In the TBF, given a frame of discernment $\Phi = \{\mathcal{C}, \overline{\mathcal{C}}\}$, where $\mathcal{C}$ means that the object is a cluster center while $\overline{\mathcal{C}}$ denotes that the object is not a cluster center, we have $2^{\Phi} = \{\emptyset, \{\mathcal{C}\}, \{\overline{\mathcal{C}}\}, \{\mathcal{C}, \overline{\mathcal{C}}\}\}$. The meta-cluster $\{\mathcal{C}, \overline{\mathcal{C}}\}$ represents total ignorance (unknown), that is, we do not know



(a) The 7-class dataset.

(b) EVCLUS (R_e = 24.43%, R_i=8.97%, CRI = 0.3736).

(c) ECM (R_e = 24.40%, R_i=17.28%, CRI = 0.8479).

(d) CCM (R_e = 31.48%, R_i=5.59%, CRI = 0.8300).

(e) BPEC (R_e = 16.87%, R_i=11.80%, CRI = 0.5773).

(f) BSC (R_e = 1.41%, R_i=3.01%, CRI = 0.8512).

Fig. 10. The clustering results of different methods on the 7-class dataset.

whether the object is a cluster center or not. Here we take the subset $\Theta = \{\{\mathcal{C}\}, \{\mathcal{C}, \overline{\mathcal{C}}\}\}$, as a new frame of discernment to describe the object as the cluster center ($\mathcal{C}$) or unknown ($\overline{\mathcal{C}}$). For short, we define $\mathcal{U} = \{\mathcal{C}, \overline{\mathcal{C}}\}$. This definition includes both $\{\mathcal{C}\}$ and $\{\overline{\mathcal{C}}\}$, and can obtain Eq. (9) based on the DS rule using Eqs. (7)-(8) quickly, as shown in the proof process.

Furthermore, we provide a detailed derivation and proof process about the degree of belief $Bel_i(A)$, which corresponds to the process of Eq. (8) to Eq. (9). Here $\mathcal{U}$ means total unknown, which means that the object may or may not be a cluster center, *i.e.* $\mathcal{C} \cap \mathcal{U} = \mathcal{C}$. For the object $\mathbf{x}_i$, its k-th neighbor $\mathbf{x}_{ik}$ provides a piece of evidence $m_{ik}(\mathcal{C})$ that supports it as a cluster center. Thus, the mass function $m_{ik}(\mathcal{U})$ on $\mathcal{C}$ is defined as $m_{ik}(\mathcal{U}) = 1 - m_{ik}(\mathcal{C})$. Thus, the degree of belief $Bel_i(\mathcal{C})$ that the object $\mathbf{x}_i$ is a center is regarded as the opposite mass of simultaneous $\mathcal{K}_1$ unknown events (*i.e.* $m_{ik}(\mathcal{U})$). Thereby, we can obtain the following proposition.

Proposition 1: The degree of belief $Bel_i(\mathcal{C})$ is defined mathematically by:

$$\begin{aligned} Bel_i(\mathcal{C}) &= 1 - \prod_{k=1}^{K_1} \left(m_{ik}(\mathcal{U})\right) \\ &= 1 - \prod_{k=1}^{\mathcal{K}_1} \left(1 - m_{ik}(\mathcal{C})\right) \end{aligned} \tag{17}$$

Next, we prove $Bel_i(\mathcal{C}) = m_i(\mathcal{C})$.

Proof: The mass function $m_i(\mathcal{U})$ of the object $\mathbf{x}_i$ as a total unknown ($\mathcal{U}$) is given by:

$$\begin{aligned} m_i(\mathcal{U}) &= \bigoplus_{k\in[1,\mathcal{K}_1]} m_{ik}(\mathcal{U}) \\ &= m_{i1}(\mathcal{U}) \bigoplus m_{i2}(\mathcal{U}) \bigoplus \cdots \bigoplus m_{i\mathcal{K}_1}(\mathcal{U}) \\ &= m_{i1}(\mathcal{U}) \times m_{i2}(\mathcal{U}) \times \cdots \times m_{i\mathcal{K}_1}(\mathcal{U}) \\ &= \prod_{k=1}^{\mathcal{K}_1} \left(m_{ik}(\mathcal{U})\right) \\ &= \prod_{k=1}^{\mathcal{K}_1} \left(1 - m_{ik}(\mathcal{C})\right) \end{aligned} \tag{18}$$

Hence, we can obtain the mass function $m_i(\mathcal{C})$ that the object $\mathbf{x}_i$ as a cluster center by taking the opposite of $m_i(\mathcal{U})$,

and denoted as:

$$m_i(\mathcal{C}) = 1 - \prod_{k=1}^{\mathcal{K}_1} (1 - m_{ik}(\mathcal{C})) \tag{19}$$

Now, we can find that $Bel_i(\mathcal{C}) = m_i(\mathcal{C})$, of course, and $Bel_i(\mathcal{U}) = m_i(\mathcal{U})$, which completes the proof.