M3W: Multistep Three-Way Clustering

Mingjing Du©, *Member, IEEE*, Jingqi Zhao, Jiarui Sun, and Yongquan Dong©

Abstract—Three-way clustering has been an active research topic in the field of cluster analysis in recent years. Some efforts are focused on the technique due to its feasibility and rationality. We observe, however, that the existing three-way clustering algorithms struggle to obtain more information and limit the fault tolerance excessively. Moreover, although the one-step three-way allocation based on a pair of fixed, global thresholds is the most straightforward way to generate the three-way cluster representations, the clusters derived from a pair of global thresholds cannot exactly reveal the inherent clustering structure of the dataset, and the threshold values are often difficult to determine beforehand. Inspired by sequential three-way decisions, we propose an algorithm, called multistep three-way clustering (M3W), to address these issues. Specifically, we first use a progressive erosion strategy to construct a multilevel structure of data, so that lower levels (or external layers) can gather more available information from higher levels (or internal layers). Then, we further propose a multistep three-way allocation strategy, which sufficiently considers the neighborhood information of every eroded instance. We use the allocation strategy in combination with the multilevel structure to ensure that more information is gradually obtained to increase the probability of being assigned correctly, capturing adaptively the inherent clustering structure of the dataset. The proposed algorithm is compared with eight competitors using 18 benchmark datasets. Experimental results show that M3W achieves superior performance, verifying its advantages and effectiveness.

Index Terms—Clustering, fuzzy-rough set theory, three-way decision theory, uncertain data analysis.

I. INTRODUCTION

AS THE most famous unsupervised learning tool, clustering technology has been developing for more than 60 years. Clustering is the tool for categorizing unlabeled data instances in such a way that data instances in the same group are more similar to each other than to those in other groups. It is widely used in various areas, including anomaly detection [1], [2], image segmentation [3], and so on [4].

Most clustering algorithms are developed under the assumption that an instance belongs to at most one group. In other words, these methods define two types of membership relations between an instance and a group, including belong-to (i.e., the instance belongs to the group) and not belong-to (i.e., the instance does not belong to the group). For ease of description, we refer to these algorithms as two-way clustering algorithms. However, in many practical situations, there are some overlapping regions between different clusters [5]. It is difficult to assign instances in the overlapping region to exactly one group. Furthermore, some applications require that an instance can be assigned to two or more groups. A news report, for example, may belong to both "sports" and "culture" in the news topic summary. It is obvious that two-way clustering is not suitable for such applications.

To address the issues above, Yu and Wang [6] propose a framework of three-way clustering, which is derived from three-way decisions [7], [8].

A. Three-Way Decision and Three-Way Clustering

A theory of three-way decision concerns problem-solving and information processing based on a particular way of human thinking known as triadic thinking [8]. According to this basic idea, a three-way decision theory is to divide a universal set into three disjoint regions and to make three types of decisions for achieving the desired outcome, accordingly. The theory is proposed originally to provide a sound semantical interpretation of decision-theoretic rough sets [7]. The subsequent studies focus on a more general sense of three-way decision that goes far beyond rough set theory. The three-way decision becomes a paradigm of thinking and information processing based on triadic patterns [8]. A more comprehensive theory includes sequential three-way decision [9], [10], statistical three-way decision [11], trisecting–acting–outcome model [8], and so on. These studies foster a number of newly emerged topics, for example, three-way attribute reduction [12], three-way clustering [6], three-way conflict analysis [13], three-way classification [14], and many more.

According to triadic thinking of three-way decision, three-way clustering defines three types of membership relations between an instance and a cluster, including belong-to certainly (i.e., the instance belongs to the cluster certainly), belong-to partially (the instance may be part of the cluster and may potentially belong to other clusters), and not belong-to certainly (i.e., the instance does not belong to the cluster certainly). Based on these membership relations, a dataset can be divided into three regions with respect to a cluster: positive region, boundary region, and negative region. The positive region of a cluster contains the data instance that are definitely a part of the cluster. The boundary region contains border
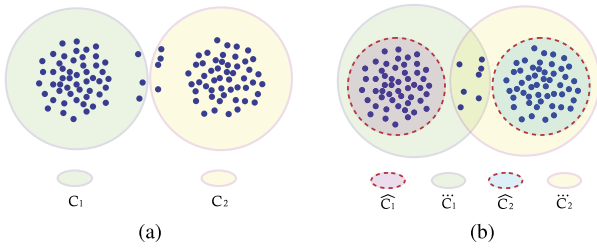
Fig. 1. Cluster representations of (a) two-way clustering and (b) three-way clustering.

data instances in the cluster that may be a part of the cluster but may also belong to other clusters. The negative region of a cluster is the collection of data instances not likely to belong to it. In three-way clustering, a cluster is represented by an interval set (a pair of nested sets called lower bound and upper bound, respectively). Fig. 1 [15] illustrates simple examples of two-way and three-way cluster representations. As shown in Fig. 1(a), two-way clustering assigns each point to a cluster ($\mathbf{C}_1$ or $\mathbf{C}_2$). In Fig. 1(b), $\widehat{\mathbf{C}}_i$ and $\overset{...}{\mathbf{C}}_i$ denote the positive region and the boundary region of $\mathbf{C}_i$, where $i = 1, 2$ (see Section III). In contrast to two-way representations of clusters, where every cluster is depicted by a set of instances, three-way clusters with interval sets can distinguish the intension and extension of each concept, contributing to better processing the data in the overlapping region, as shown in Fig. 1(b).

B. Motivation and Contribution

Recently, a series of three-way clustering approaches have been proposed and studied. Although these methods are significantly better than two-way ones in terms of rationality,[1] they have some shortcomings, which may hinder their widespread applications.

1) Existing three-way clustering algorithms adopt an allocation strategy or clustering strategy based on the one-step (single) three-way decision-making process. These algorithms do not struggle to obtain more available information and excessively limit fault tolerance. When there is no detailed information available, some data may be assigned incorrectly. Even worse, in three-way extensions of density-based clustering algorithms, a wrongly assigned instance can cause greater issues in subsequent allocations, leading to unsatisfactory performance.

2) Most existing three-way clustering methods adopt an evaluation function and a pair of partition thresholds to obtain the three-way representations of clusters. Generally, these algorithms apply a pair of fixed, global thresholds across the entire dataset. However, the threshold setting is only applicable to the clusters with the same characteristics, not to the clusters with different distributions and different densities. Fundamentally, these algorithms cannot sufficiently account for both the

inherent characteristics of every cluster and the positions of each instance. As a result, these algorithms fail to recognize clusters with different distribution patterns.

Inspired by the sequential three-way decision [9], [10], we propose a multistep three-way clustering algorithm, called M3W. Sequential three-way decisions are also known as multistep three-way decisions. The sequential three-way decision theory is initially developed to deal with a cost-sensitive decision-making problem based on multiple levels of granularity. The multilevel structure may take the form of multiple representations of data or multiple descriptions of problems. A multilevel structure creates a partial ordering relation where the finer granularities are represented at lower levels, and the coarser granularities are represented at higher levels. At a higher level, we may make the acceptance or rejection decisions for certain data with sufficient information and do not make decisions immediately for some data with incomplete information. By moving to a lower level, these remaining data can be investigated again, and a proper judgment can be given, once we have enough information about them.

As with density-based clustering, the proposed algorithm is based on the assumption that the core regions of clusters should be separated implicitly by the boundary regions of clusters [16], [17]. Based on this assumption, we use density information of data to construct a multilevel structure of data, where the higher levels (or the internal layers) with a higher density are closer to core regions of clusters, and the lower levels (or the external layers) with a lower density are closer to boundary regions of clusters. By introducing the idea of sequential three-way decisions, we design a multistep clustering process. After the multilevel structure is created, all clusters can be formed using the clustering process from the highest level down.

Specifically, we first generate a multilevel structure of data using a progressive erosion strategy [16], [17]. In the multilevel structure, the data at the highest level (i.e., the data from the core regions) can be better separated and clustered more easily. In contrast, the data at a lower level may be located at the boundary regions of clusters, even the overlapping regions of clusters. Then, we use an adaptive scheme that determines automatically the distance threshold of every instance. With the help of the adaptive scheme, a connectivity-based approach is used to cluster the core data (i.e., data at the highest level). As a result, it is possible to detect clusters with a variety of distributions. Finally, the data at lower levels ought to get adequate attention, since they can be potentially misclassified. To assign the data at the lower levels to the correct clusters, we try to consider the cluster assignments the data at the higher levels (i.e., more detailed information). We use a "two-stage" allocation scheme based on the deferred decision[2] to assign the data at the lower levels (or the external layers). The proposed clustering can gradually construct the core region and boundary region of each cluster, following the levels from high to low.

---

[1] A three-way cluster with interval sets shows that not only does an instance certainly belong to the cluster but also that an instance might belong to the cluster intuitively.

[2] In three-way decision, the deferred decision is viewed as the third decision-making behavior. Deferred decision making is an option when the available information is insufficient, or the evidence is not strong enough to support an acceptance or rejection.

The contributions of this work are summarized as follows.

1) We propose a novel three-way clustering, called M3W, that combines the strengths of progressive erosion strategy and sequential three-way decision. Different from most existing three-way clustering methods, M3W is designed as an iteration process, which can handle clusters with complex distributions and unclear boundaries better.

2) As part of the proposed algorithm, we use the progressive erosion strategy as a means of generating a multilevel structure of data, so that the initial core regions of nearby clusters can be separated clearly. This helps M3W detect clusters with overlapping regions or unclear boundaries.

3) With the aid of the generated multilevel structure, we craft a multistep clustering process that utilizes the idea of the sequential three-way decision.[3] The process uses fully the neighborhood information of every instance to determine their class assignments level by level. All these make it possible to capture the inherent clustering structure of the dataset with different characteristics.

Here is an outline of the rest of this article. We review the studies that are related to our work in Section II. Section III briefly introduces the background knowledge, notations, and definitions. The details of M3W are introduced in Section IV. Section V presents the experiments on synthetic and real datasets. Finally, we summarize the concluding remarks and future work in Section VI.

II. Related Work

In recent decades, soft clustering algorithms have been proposed and applied successfully. The fuzzy C-means (FCM) [18] is probably still the most widely used algorithm for soft clustering. By introducing rough set theory, Lingras and West [19] propose the rough K-means algorithm (RKM), which uses interval sets to represent clusters with vague and imprecise boundaries. In an effort to enrich this field further, Yu *et al.* [6] propose three-way clustering by incorporating three-way decisions into clustering. Due to the convenience of storing and computing, three-way clustering algorithms have been extended to accommodate different scenarios, and a series of algorithms have been developed, such as consensus clustering [20], [21], [22], incremental three-way clustering [23], three-way clustering for network data [24], three-way clustering for incomplete data [25], and multiview three-way clustering [26], [27]. Here, we review and analyze the most closely related research to our algorithm.

Wang and Yao [28] introduce the erosion and dilation ideas from mathematical morphology into K-means and propose a general framework of three-way clustering based on a contraction-and-expansion strategy, called CE3. Yu *et al.* [29] propose a three-way clustering method based on DBSCAN (3W-DBSCAN). With the use of a type function and a pair of density thresholds, this algorithm can convert a two-way

---

[3]Its main idea is that we do not make a decision immediately when the information is insufficient or incomplete until we have enough information about these problems to make the right decision.

density-based cluster into a three-way density-based cluster. However, the clustering results obtained by the two approaches (CE3 and 3W-DBSCAN) are heavily dependent on the performance of the corresponding original algorithms (K-means and DBSCAN). Different from CE3, TWKM [15] is another version of three-way K-means, which adopts perturbation analysis to separate the core regions from the supports. Zhang [30] proposes a three-way c-means algorithm (called TCM) by integrating the three-way weight and three-way assignment. 3WDPET [31] is a three-way density peak clustering method that combines evidence theory and density peak clustering. However, the four algorithms (CE3, TWKM, TCM, and 3WDPET) need a specification of the number of clusters in advance.

Clusters are usually represented as three-way representations by means of a pair of partition thresholds. In practice, however, it is very difficult to tune these threshold values. Several studies are conducted in response to this issue. 3WC-OR [32] uses the genetic algorithms to determine automatically the partition thresholds. In the same year, Yu *et al.* [33] propose an efficient three-way clustering algorithm based on the idea of universal gravitation, which can automatically adjust partition thresholds. A threshold selection approach is proposed by Jia *et al.* [34] to improve three-way clustering. However, unlike our algorithm, where the three-way clusters are formed gradually by an adaptive multistep allocation scheme, these three algorithms are based on an automatic threshold selection approach, where the clustering results are driven by the estimated thresholds.

III. Background and Related Studies

A. Notations

Consider a set of N instances $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, where $\mathbf{x}_i \in \mathbb{R}^M, 1 \leqslant i \leqslant N$. For an erosion process of L levels, $\mathbf{X}^{(l)}$ denotes a set of the remaining uneroded data at the lth level, and $\mathbf{X}_E^{(l)}$ denotes a set of eroded data at the lth level. The set of uneroded data at the higher level is given by $\mathbf{X}^{(l+1)} = \mathbf{X}^{(l)} \setminus \mathbf{X}_E^{(l)}$. Before the erosion process is not performed, $\mathbf{X}^{(1)} = \mathbf{X}$. $\mathbf{X}_B^{(l)}$ denotes a set of eroded data at the first l levels, as follows:

$$\mathbf{X}_B^{(l)} = \mathbf{X}_E^{(1)} \cup \cdots \cup \mathbf{X}_E^{(l)} \qquad (1)$$

where $\|\mathbf{x}_i - \mathbf{x}_j\|$ denotes the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. $N_k^{(l)}(\mathbf{x}_i)$ denotes the kth nearest neighbor of $\mathbf{x}_i$ at the lth level. $\mathrm{kNN}^{(l)}(\mathbf{x}_i)$ denotes a set of k nearest neighbors (k-NN) of an instance $\mathbf{x}_i$ at the lth level. For each instance $\mathbf{x}_j \in \mathrm{kNN}^{(l)}(\mathbf{x}_i)$, we have $\mathbf{x}_j \in \mathbf{X}^{(l)}$. Similarly, a set of reverse k nearest neighbors (reverse k-NN) of an instance $\mathbf{x}_i$ at the lth level can be denoted by $RkNN^{(l)}(\mathbf{x}_i)$.

$\mathrm{kNN}_B^{(l)}(\mathbf{x}_i)$ denotes a set of k nearest eroded neighbors (i.e., each eroded neighbor $\mathbf{x}_j \in \mathbf{X}_B^{(l-1)}$) of an instance $\mathbf{x}_i$ in $\mathbf{X}^{(l)}$.

For sake of readability, the abovementioned notations are summarized in Table I.

B. Three-Way Representation

In terms of a three-way representation, a dataset can be divided into three regions with respect to a cluster $\mathbf{C}_i$: positive

TABLE I
NOTATIONS IN M3W CLUSTERING

Notation	Definition
N	Number of all data instances
M	Number of attributes in each instance
k	Number of nearest neighbors
K	Number of clusters
L	Number of levels
$\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$	Data set of all data instances
$\mathbf{X}^{(l)}$	Set of remain data at l-th level
$\mathbf{X}_E^{(l)}$	Set of eroded data at l-th level
$\mathbf{X}_B^{(l)}$	Set of eroded data at the first l levels
$\|\mathbf{x}_i - \mathbf{x}_j\|$	Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$
$N_k^{(l)}(\mathbf{x}_i)$	k-th nearest neighbor of an instance $\mathbf{x}_i$ at l-th level
$kNN^{(l)}(\mathbf{x}_i)$	Set of k nearest neighbors of an instance $\mathbf{x}_i$ at l-th level
$RkNN^{(l)}(\mathbf{x}_i)$	Set of reverse k nearest neighbors of an instance $\mathbf{x}_i$ at l-th level
$kNN_B^{(l)}(\mathbf{x}_i)$	Set of of k nearest eroded neighbors of an instance $\mathbf{x}_i$ at l-th level

region (or called core region, $\widehat{\mathbf{C}}_i$), boundary region ($\overset{...}{\mathbf{C}}_i$), and negative region ($\widetilde{\mathbf{C}}_i$). The data instances in $\widehat{\mathbf{C}}_i$ definitely belong to $\mathbf{C}_i$. The data instances in $\overset{...}{\mathbf{C}}_i$ possibly belong to $\mathbf{C}_i$ and possibly belong to other clusters. The data instances in $\widetilde{\mathbf{C}}_i$ definitely do not belong to $\mathbf{C}_i$.

In three-way clustering, a cluster is represented by a pair of nested sets [6], as follows:

$$\mathbf{C}_i = (\underline{\mathbf{C}_i}, \overline{\mathbf{C}_i}) = \left(\widehat{\mathbf{C}}_i, \widehat{\mathbf{C}}_i \cup \overset{...}{\mathbf{C}}_i\right) \qquad (2)$$

where $\underline{\mathbf{C}_i}$ is the lower bound of $\mathbf{C}_i$, and $\overline{\mathbf{C}_i}$ is the upper bound of $\mathbf{C}_i$. Obviously, $\underline{\mathbf{C}_i} = \widehat{\mathbf{C}}_i$, $\overline{\mathbf{C}_i} = \widehat{\mathbf{C}}_i \cup \overset{...}{\mathbf{C}}_i$, and $\underline{\mathbf{C}_i} \subseteq \overline{\mathbf{C}_i} \subseteq \mathbf{X}$.

Therefore, the three-way clustering results can be described as follows:

$$\mathcal{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_K\} = \left\{(\underline{\mathbf{C}_i}, \overline{\mathbf{C}_i}) | 1 \leqslant i \leqslant K\right\}. \qquad (3)$$

A three-way cluster satisfies the following properties [6].

Property 1 (Non-Emptiness): The lower bound of each cluster cannot be empty. $\underline{\mathbf{C}_i} \neq \emptyset$, $1 \leqslant i \leqslant K$.

Property 2 (Universe): The union of the upper bounds of all clusters is the universe (the whole dataset). $\overline{\mathbf{C}_1} \cup \cdots \cup \overline{\mathbf{C}_K} = \mathbf{X}$.

Property 3 (Mutual Exclusion): The intersection of lower bounds of any two clusters is empty. $\underline{\mathbf{C}_i} \cap \underline{\mathbf{C}_j} = \emptyset$, $1 \leq i, j \leq K$ and $i \neq j$.

Property 1 implies that a cluster cannot be empty. It ensures that every cluster is meaningful. Properties 2 and 3 state that an instance must belong to one or more clusters.

IV. M3W METHOD

In this section, we first explain each phase of M3W, whose architecture is shown in Fig. 2. Then, we analyze its computational complexity.

A. Dynamic Density Estimation

Kernel density estimation is a widely used density measure, and the kernel density estimator of $\mathbf{x}_i$ is given by the following equation:

$$\rho_i = \sum_{\mathbf{x}_j \in \mathbf{V}} \kappa\left(\frac{\|\mathbf{x}_j - \mathbf{x}_i\|}{h}\right) \qquad (4)$$

where $\kappa(\cdot)$ is a nonnegative and monotonically decreasing function, whereas h is a term for bandwidth used to control scale. Variables $\mathbf{x}_i$ and $\mathbf{x}_j$ refer to a test instance and a sample, respectively. $\mathbf{V}$ is a set of samples $\mathbf{x}_j$.

As for the kernel function, we use the most widely used kernel–Gaussian kernel function because of its smoothness.

If a kernel function can be seen as a filter, the bandwidth h is a weighted term in the filter. Some work takes the bandwidth h as an input parameter. This scheme, however, poses two problems. The first problem is having to specify the parameter by the user, thereby reducing its usability. The second problem is that a filter with a global weight may imply that data instances in low-density regions (the distance between them is large) are given a relatively low weight, whereas data instances in high-density regions are given a relatively high weight. Different regions of the kernel are expected to have different bandwidths. A sample and common adaptive scheme is used [35]: h_j is set to be equal to the distance between $\mathbf{x}_j$ to its kth nearest neighbor. According to the scheme, the bandwidth varies with the location of the sample.

To discover the distribution information of data, we utilize reverse k nearest neighbors (Rk-NN). Unlike traditional adaptive Gaussian kernel density estimation-based, our method employs a dynamic strategy, in which the density of the remaining data can be re-estimated at each level

$$\rho_i^{(l)} = \sum_{\mathbf{x}_j \in RkNN^{(l)}(\mathbf{x}_i)} \exp\left(-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{\|\mathbf{x}_j - N_k^{(l)}(\mathbf{x}_j)\|^2}\right). \qquad (5)$$

B. Progressive Erosion Strategy

Our algorithm is based on the observation that boundaries between different clusters tend to be misclassified, especially when they are not obvious. According to the observation, we can conclude that the cores of clusters may be better clustered than their boundaries. In this work, we use a progressive erosion strategy to discover the "initial" cores of the latent clusters. Different from traditional approaches, which adopt a fixed, global threshold to define directly the cores of the clusters, an erosion strategy can erode iteratively the boundaries between clusters to draw automatically the "initial" cores. Due to the introduction of the dynamic density estimation in the progressive erosion strategy, the data instances on the external part of the cluster are eroded earlier than those on the inner part, despite being at the same density level (based on the non-dynamic density estimation).

Each level of erosion involves three steps. The first step is to estimate the density for each instance at the current level by executing (5). The second step is to sort the data according to density values. Third, some of these data are
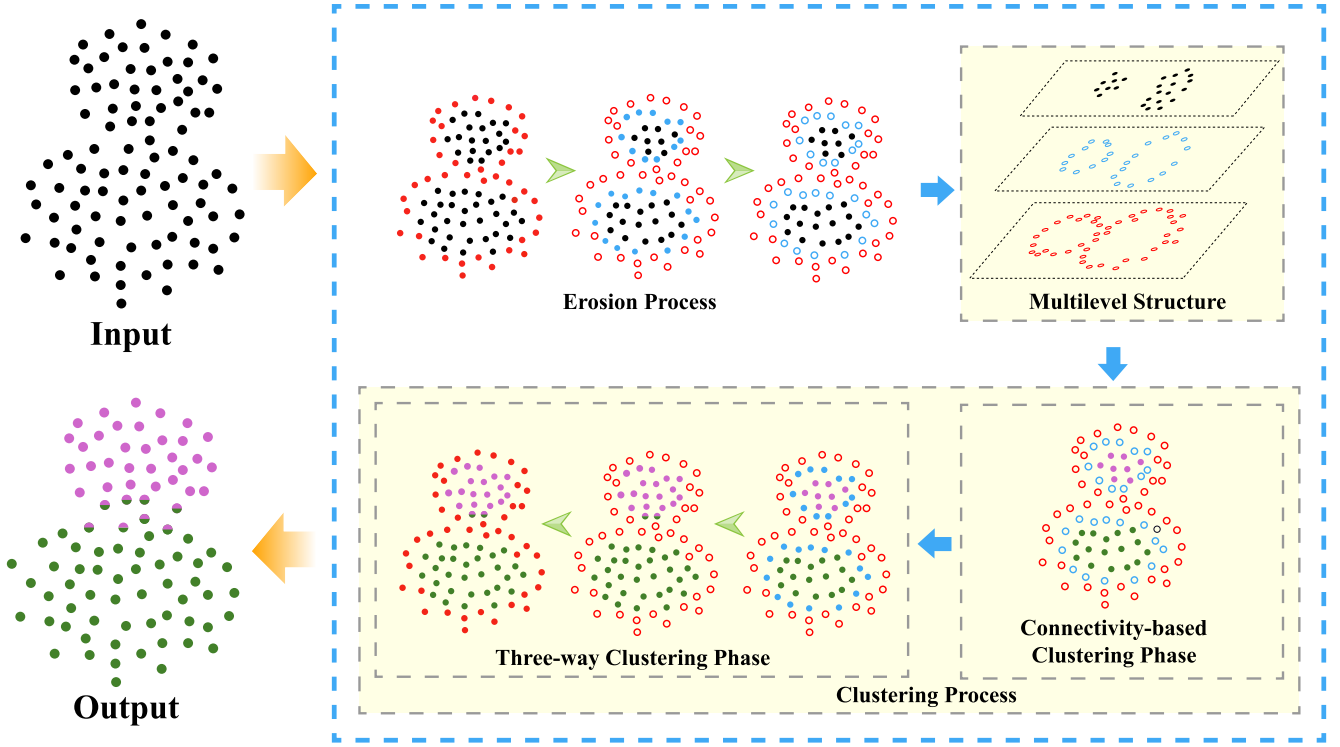
Fig. 2. Architecture of M3W.

deemed boundaries and are eroded. Formally, the eroded data whose density values are smaller than a cutoff value can be given by the following equation:

$$\mathbf{X}_E^{(l)} = \left\{ \mathbf{x}_i \mid \rho_i^{(l)} \leqslant \rho_c^{(l)} \right\}. \tag{6}$$

Similar to [36] and [37], a series of cutoff values can be given indirectly by a percentile. Specifically, the data instances with density values in the tenth percentile or lower are considered eroded data. In other words, $\rho_c^{(l)}$ is set to remain 90% of $\mathbf{X}^{(l)}$ whose density values are relatively higher at each iteration. At the highest level, the set of uneroded data is as follows:

$$\mathbf{X}^{(l+1)} = \mathbf{X}^{(l)} \setminus \mathbf{X}_E^{(l)}. \tag{7}$$

The highly separable data instances remain after the erosion process. The set of these final remaining data is denoted by $\mathbf{X}^{(L+1)}$.

While revealing the "initial" cores of the latent clusters, the progressive erosion process produces a multilevel structure of the data, i.e., $\mathbf{X}_E^{(1)}, \ldots, \mathbf{X}_E^{(L)}, \mathbf{X}^{(L+1)}$. With the aid of the generated multilevel structure, we can easily reconstruct latent clusters by using the technique we describe next.

C. Multistep Clustering Process

To construct the three-way clusters, we use a multistep approach based on sequential three-way decisions. The proposed multistep clustering process mainly consists of two clustering phases: the data instances in the set $\mathbf{X}^{(L+1)}$ are clustered by using a connectivity-based approach; the eroded data instances are assembled in the order (the reverse order

of the erosion, i.e., $\mathbf{X}_E^{(L)}, \ldots, \mathbf{X}_E^{(2)}, \mathbf{X}_E^{(1)}$) by using a three-way approach.

1) Connectivity-Based Approach: Inspired by DBSCAN [38] and HDBSCAN [39], we define a connectivity-based approach to cluster the final remaining data, $\mathbf{X}^{(L+1)}$. Before presenting the details of this connectivity-based approach, we first formalize some concepts in the following.

Definition 1 (Core Distance): The core distance $d_{\text{core}}(\mathbf{x}_i)$ of every instance $\mathbf{x}_i \in \mathbf{X}$ with respect to k is the distance from $\mathbf{x}_i$ to its kth nearest neighbor.

Definition 2 (Adaptive Distance Threshold): The adaptive distance threshold of every instance $\mathbf{x}_i$ in the lth level is defined as follows:

$$\varepsilon_{\text{core}}^{(l)}(\mathbf{x}_i)$$
$$= \begin{cases} \text{Mean}(D_{\text{core}}) + \text{Std}(D_{\text{core}}), & l = 1 \\ \min\left\{ \sum_{\mathbf{x}_j \in \text{kNN}_B^{(l)}(\mathbf{x}_j)} \frac{\gamma}{k} \|\mathbf{x}_i - \mathbf{x}_j\|, \varepsilon_{\text{core}}^{(l-1)} \right\}, & 2 \leq l \leq L+1 \end{cases} \tag{8}$$

where $\text{kNN}_B^{(l)}(\mathbf{x}_j)$ denotes a set of k nearest eroded neighbors of an instance $\mathbf{x}_i$ in $\mathbf{X}^{(l)}$ (see Section II-A). γ is a constant factor. We empirically determine its value as 3.

Definition 3 (Mutual Reachability Distance Threshold): The mutual reachability distance threshold between two data instances $\mathbf{x}_i$ and $\mathbf{x}_j$ in $\mathbf{X}^{(L+1)}$ is defined as $\varepsilon_{mr}(\mathbf{x}_i, \mathbf{x}_j) = \max\{\varepsilon_{\text{core}}^{(L+1)}(\mathbf{x}_i), \varepsilon_{\text{core}}^{(L+1)}(\mathbf{x}_j)\}$.

Definition 4 (Mutual Reachability): Two data instances $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}^{(L+1)}$ are mutual reachable, if there is a series of

data instances $\mathbf{x}_{p_1}, \ldots, \mathbf{x}_{p_m}$ with $p_1 = i$ and $p_m = j$, where the distance between any two adjacent data instances $\mathbf{x}_{p_s}, \mathbf{x}_{p_{s+1}}$ is less than or equal to the mutual reachability distance threshold between them, i.e., $\|\mathbf{x}_{p_s} - \mathbf{x}_{p_{s+1}}\| \leq \varepsilon_{mr}(\mathbf{x}_{p_s}, \mathbf{x}_{p_{s+1}})$.

The "initial" cores of the latent clusters are formed by merging all data instances in $\mathbf{X}^{(L+1)}$ that are mutually reachable. The clustering result at the $(L + 1)$th level can be denoted by $\mathcal{C}^{(L+1)} = \{\mathbf{C}_1^{(L+1)}, \ldots, \mathbf{C}_K^{(L+1)}\} = \{(\underline{\mathbf{C}_1^{(L+1)}}, \overline{\mathbf{C}_1^{(L+1)}}), \ldots, (\underline{\mathbf{C}_K^{(L+1)}}, \overline{\mathbf{C}_K^{(L+1)}})\}.$[4]

Note that all data instances in $\mathbf{X}^{(L+1)}$ are assigned to the core regions of the clusters; thus, the clustering result at the $(L + 1)$th level can also be denoted by $\mathcal{C}^{(L+1)} = \{\underline{\mathbf{C}_1^{(L+1)}}, \ldots, \underline{\mathbf{C}_K^{(L+1)}}\}$.

2) Three-Way Approach: In subsequent steps, a three-way approach can sequentially process $\mathbf{X}_E^{(L)}, \ldots, \mathbf{X}_E^{(2)}, \mathbf{X}_E^{(1)}$. The process starts from the data at the Lth level, i.e., $\mathbf{X}_E^{(L)}$. At each level, a "two-stage" allocation scheme based on the deferred decision (i.e., one-step three-way allocation approach) utilizes the neighbor information to cluster data at the current level. The allocation scheme consists of two three-way decision rules: 1) assign the data at the current level to the core region or "candidate" boundary region of the corresponding cluster and 2) reassign the data in the "candidate" boundary regions to the core region or boundary region of the corresponding cluster. For convenience, we give some basic definitions.

$\mathbf{X}_{\text{core}}^{(l)}$ denotes the union of cores of all current clusters at the lth level, i.e., $\mathbf{X}_{\text{core}}^{(l)} = \underline{\mathbf{C}_1^{(l)}} \cup \cdots \cup \underline{\mathbf{C}_K^{(l)}}$.

Definition 5 (Core Neighbors): For every instance $\mathbf{x}_i \in \mathbf{X}_E^{(l)}$, the set of core k nearest neighbors of $\mathbf{x}_i$, $CN^{(l)}(\mathbf{x}_i)$ is its k nearest neighbors from $\mathbf{X}_{\text{core}}^{(l+1)}$.

An instance $\mathbf{x}_i \in \mathbf{X}_E^{(l)}$ is assigned to the core region or boundary region of its corresponding cluster according to the probability of being a member of that cluster.

The probability is defined as follows:

$$P\left(\mathbf{C}_r^{(l)}|\mathbf{x}_i\right) = \frac{\left|\left\{\mathbf{x}_j|\mathbf{x}_j \in CN^{(l)}(\mathbf{x}_i)\right\} \cap \left\{\mathbf{x}_p|\mathbf{x}_p \in \underline{\mathbf{C}_r^{(l)}}\right\}\right|}{\left|CN^{(l)}(\mathbf{x}_i)\right|} \quad (9)$$

where $|\mathbf{X}|$ denotes the cardinality of the set $\mathbf{X}$.

In the frequency-based membership, for every instance $\mathbf{x}_i \in \mathbf{X}_E^{(l)}$, $P(\mathbf{C}_r^{(l)}|\mathbf{x}_i)$ provides the percentage of its core neighbors who belong to a specific cluster.

For every unlabeled instance $\mathbf{x}_i \in \mathbf{X}_E^{(l)}$, its probability distribution $P(\mathcal{C}^{(l)}|\mathbf{x}_i) = \{P(\mathbf{C}_1^{(l)}|\mathbf{x}_i), \ldots, P(\mathbf{C}_r^{(l)}|\mathbf{x}_i), \ldots, P(\mathbf{C}_K^{(l)}|\mathbf{x}_i)\}$ can be regarded as a vector of probabilities. The length K is the number of clusters. Obviously, the rth entry of the vector indicates the probability that the instance belongs to the rth cluster.

We use the probability vector to define a "two-stage" allocation scheme based on the deferred decision (i.e., two three-way decision rules). Next, we present the definitions of Rules 1 and 2.

Rule 1 is given as follows.

If all core neighbors of $\mathbf{x}_i$ belong only to the cluster $\mathbf{C}_r$, i.e., if $P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) = 1$, the instance $\mathbf{x}_i$ is assigned to the core region of $\mathbf{C}_r$.

If its core neighbors do not all belong to the cluster $\mathbf{C}_r$ and the probability of being a member of that cluster is higher than that of other clusters, i.e., if $0 < P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) < 1$ and $P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) = \max(P(\mathcal{C}^{(l)}|\mathbf{x}_i))$, the instance $\mathbf{x}_i$ is assigned to the "candidate" boundary region of $\mathbf{C}_r$. We use Rule 2 to further process and assign them to the proper clusters.

Rule 2 is given as follows.

For every instance $\mathbf{x}_i$ in the "candidate" boundary region of $\mathbf{C}_r$, if there is a cluster $\mathbf{C}_s, s \neq r$, such that the gap between probabilities of being in $\mathbf{C}_s$ and $\mathbf{C}_r$ is less than $(1/K)$, i.e., if $P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) - P(\mathbf{C}_s^{(l)}|\mathbf{x}_i) < (1/K)$ [31], we assign the instance $\mathbf{x}_i$ to the boundary region of $\mathbf{C}_r$ and add it into the boundary regions of clusters that satisfy the condition. Otherwise, the instance is assigned to the core region of $\mathbf{C}_r$.

During the clustering process, the data instances at lower levels that are more susceptible to being misclassified obtain gradually more available information (i.e., the cluster assignments of their neighbors). We use the "two-stage" allocation scheme to assign as many border data as possible to the correct clusters.

This "sequential" three-way process continues until all unlabeled instances at the 1th level are visited and allocated. In the clustering process, all clusters are formed gradually from the highest level down.

D. Algorithm and Complexity Analysis

In summary, the procedure of M3W algorithm is listed in Algorithm 1.

The computational complexity of M3W is dependent on the complexity of the k-NN search method. The complexity of the method is $O(kN\log N)$ for data indexed by some metric tree approaches (e.g., the R* tree and k–d tree) or $O(kN^2)$ for nonindexed data, where N is the number of the data, and k is the number of the neighbors.

In the erosion phase (from Lines 1 to 6), generating a multilevel structure requires $O(Lk\bar{N}_C \log \bar{N}_C) + O(L\bar{N}_C)$ operations, where L is the number of levels, and $\bar{N}_C$ is the average number of all $|\mathbf{X}^{(l)}|$. The complexity of the connectivity-based clustering phase in Lines 7–9 is $O(kN_C)$, where N_C is the cardinality of $\mathbf{X}^{(L+1)}$ at the $(L+1)$th level. The computational complexity of the three-way clustering phase in Lines 10–21 is $O(Lk\bar{N}_E)$, where the average number of all $|\mathbf{X}_E^{(l)}|$. So, the total time cost is $O(Lk\bar{N}_C \log \bar{N}_C + L\bar{N}_C + kN_C + Lk\bar{N}_E) \sim O(Lk\bar{N}_C \log \bar{N}_C)$. Since $L \ll N$, $k \ll N$, $\bar{N}_C \approx N$, the overall complexity of M3W is about $O(N\log N)$.

V. EXPERIMENTS

In this section, we compare the proposed M3W with eight clustering algorithms on 18 benchmark datasets.

A. Experiment Setup

Eighteen datasets are used to evaluate the performance of the proposed algorithm. The datasets are divided into two

---

[4]More generally, the clustering result at the lth level can be denoted by $\mathcal{C}^{(l)} = \{\mathbf{C}_1^{(l)}, \ldots, \mathbf{C}_K^{(l)}\} = \{(\underline{\mathbf{C}_1^{(l)}}, \overline{\mathbf{C}_1^{(l)}}), \ldots, (\underline{\mathbf{C}_K^{(l)}}, \overline{\mathbf{C}_K^{(l)}})\}$.

Algorithm 1 Proposed M3W Algorithm

Input: A set of data, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$
The number of neighbors, k
The number of levels, L

Output: The clustering result,
$\mathcal{C} = \{(\underline{\mathbf{C}_1}, \overline{\mathbf{C}_1}), \ldots, (\underline{\mathbf{C}_K}, \overline{\mathbf{C}_K})\}$

// produce a multilevel structure in the erosion phase

1   $\mathbf{X}^{(1)} \leftarrow \mathbf{X}$;

2   **for** $l = 1; l \leq L; l + +$ **do**

3   **foreach** $\mathbf{x}_i$ *in* $\mathbf{X}^{(l)}$ **do**

4       $\rho_i^{(l)} =$
      $\displaystyle\sum_{\mathbf{x}_j \in RkNN^{(l)}(\mathbf{x}_i)} exp(-(\|\mathbf{x}_j - \mathbf{x}_i\|^2 / \|\mathbf{x}_j - N_k^{(l)}(\mathbf{x}_j)\|^2))$;

5     $\mathbf{X}_E^{(l)} = \{\mathbf{x}_i \mid \rho_i^{(l)} \leqslant \rho_c^{(l)}\}$;

6     $\mathbf{X}^{(l+1)} = \mathbf{X}^{(l)} \setminus \mathbf{X}_E^{(l)}$;

// obtain $\mathcal{C}^{(L+1)}$ in the connectivity-based clustering phase

7   **foreach** $\mathbf{x}_i$ *in* $\mathbf{X}^{(L+1)}$ **do**

8   **foreach** $\mathbf{x}_j$ *in* $kNN^{(L+1)}(\mathbf{x}_i)$ **do**

9   **if** $\|\mathbf{x}_i - \mathbf{x}_j\| \leq \varepsilon_{mr}(\mathbf{x}_i, \mathbf{x}_j)$ **then** merge $\mathbf{x}_i$ and $\mathbf{x}_j$

// obtain $\mathcal{C}^{(l)}, 1 \leq l \leq L,$ in the three-way clustering phase

10   **for** $l = L; l \geq 1; l - -$ **do**

11     $\mathcal{C}^{(l)} \leftarrow \mathcal{C}^{(l+1)}$;

12   **foreach** $\mathbf{x}_i$ *in* $\mathbf{X}_E^{(l)}$ **do**

13   **if** $P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) = 1$ **then**

14         $\overline{\mathbf{C}_r^{(l)}} = \overline{\mathbf{C}_r^{(l)}} \cup \mathbf{x}_i$;

15   **else**

16   **if** $\exists j, \ s.t. \ P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) - P(\mathbf{C}_j^{(l)}|\mathbf{x}_i) < (1/K)$ **then**

17           $\overline{\mathbf{C}_r^{(l)}} = \overline{\mathbf{C}_r^{(l)}} \cup \mathbf{x}_i$;

18   **foreach**
          $s \in \{j|P(\mathbf{C}_r^{(l)}|\mathbf{x}_i) - P(\mathbf{C}_j^{(l)}|\mathbf{x}_i) < (1/K)\}$ **do**
          $\overline{\mathbf{C}_s^{(l)}} = \overline{\mathbf{C}_s^{(l)}} \cup \mathbf{x}_i$

19   **else**

20           $\underline{\mathbf{C}_r^{(l)}} = \underline{\mathbf{C}_r^{(l)}} \cup \mathbf{x}_i$;

21   $\mathcal{C} \leftarrow \mathcal{C}^{(1)}$;

TABLE II

DATASETS USED IN EXPERIMENTS

Data sets	#Instances	#Features	#Classes
Triangle1	1000	2	4
Triangle2	1000	2	4
S1	5000	2	15
S2	5000	2	15
T2	4000	2	6
Pathbased	300	2	3
Ds2c2sc13	588	2	13
T4	7236	2	6
Glass	214	10	6
Dermatology	366	34	6
Digits	1797	64	10
MSRA	1799	256	12
Segmentation	2310	19	7
Optdigits	5620	64	10
Statlog	6435	36	6
Pendigits	10992	16	10
Htru	17898	8	2
Shuttle	22170	9	3

The evaluation of our algorithm is conducted by comparing it with eight other clustering algorithms, including 3W-DBSCAN [29], 3W-DPET [31], CE3 [28], NEO-K-Means [40], Fuzzy C-means (FCM) [18], rough K-Means (RKM) [19], Kernel K-means (KnK-Means) [41], and spectral clustering (SC) [42]. Among them, 3W-DBSCAN, 3W-DPET, and CE3 are three state-of-the-art three-way clustering algorithms. As similar to three-way clustering, NEO-K-Means can also be used to identify overlapping regions between clusters. FCM and RKM are two benchmark algorithms for soft clustering. The KnK-Means and SC algorithms are two representative two-way clustering algorithms. Both are able to identify clusters that have nonlinear shapes.

There are two parameters in M3W, namely, the number of neighbors (k) and the number of levels (L). k is a positive integer whose values range from 5 to 30. In addition, L has a value between 2 and 12. η, ε, and k are three parameters of 3W-DBSCAN that indicate the neighborhood radius of the scaling function, the distance threshold, and the density threshold (in DBSCAN, the density threshold is called *minPts*), respectively. η and ε take values from 0.1 to 1 with a step of 0.1. The parameter k range is the same as that of M3W. The 3W-DPET and CE3 both have parameters k and K, which describe the number of neighbors and clusters, respectively. The values of k of two algorithms are selected continuously from 5 to 30. NEO-K-Means has three parameters α, β, and K, which represent the factor of overlap and the factor of nonexhaustiveness and the number of clusters, respectively. Using the parameter settings in [40], the value of α varies from -1 to 3.5 by increasing 0.5, and the parameter range of β is $\{3, 6\}$. FCM has two parameters K and m, which describe the number of clusters and the fuzzy exponent, respectively. The value of m varies from 2 to 5 by increasing 0.5. Three parameters of RKM are K, w_u, and ϵ, which represent the number of clusters, the upper approximate weight, and the threshold of the ratio, respectively. w_u ranges from 0.1 to 0.4 with a step of 0.1, and w_u ranges from 0.7 to 1 with

groups: eight synthetic datasets and ten real-world datasets. For visual convenience, all synthetic datasets used in our experiments are 2-D datasets, including Triangle1, Triangle2, S1, S2, T2, Pathbased, Ds2c2sc13, and T4. Real-world datasets include Glass, Dermatology, Digits, MSRA, Segmentation, Optdigits, Statlog, Pendigits, Htru, and Shuttle. These benchmark datasets can be found at some published benchmarks, such as UCI Machine Learning Data Repository[5] and clustering benchmark datasets.[6] Table II summarizes the detailed information of these datasets.

[5]http://archive.ics.uci.edu/ml/index.php
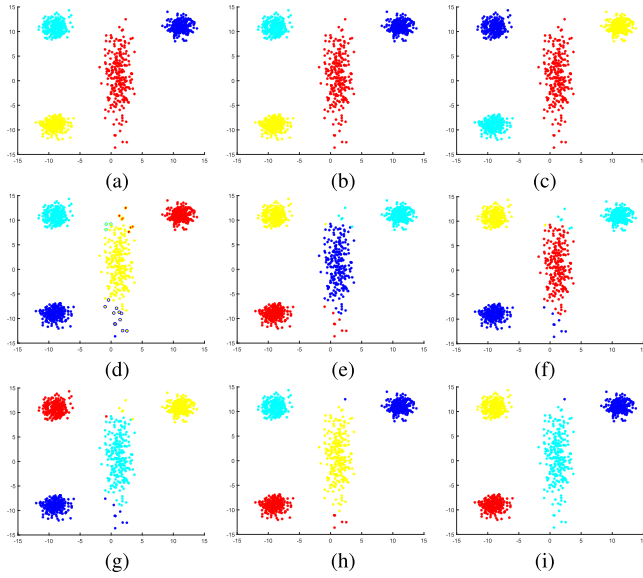[6]https://github.com/deric/clustering-benchmark

Fig. 3. Clustering results on Triangle1. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.

Fig. 4. Clustering results on Triangle2. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.

Fig. 5. Clustering results on S1. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.
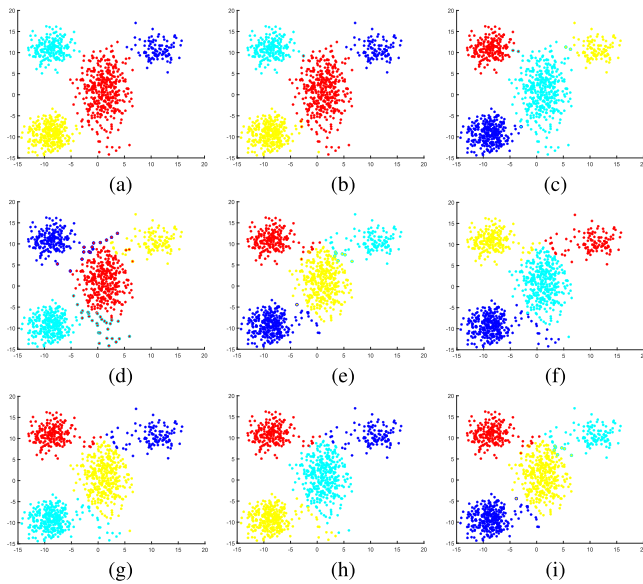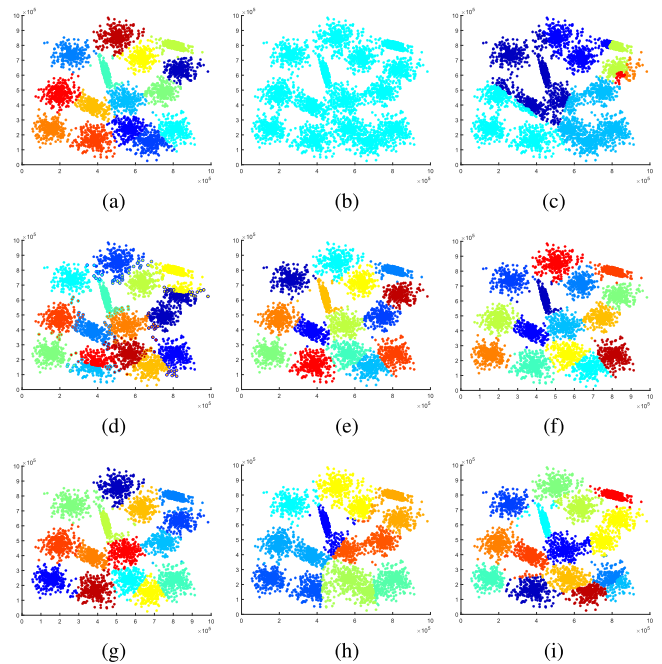

Fig. 6. Clustering results on S2. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.

a step of 0.1. In kernel K-Means, the number of clusters (K) is the only parameter. Two parameters of SC are K and σ, which represent the number of clusters and the factor of the width of the neighborhoods, respectively. σ ranges from 0.5 to 4, with a step of 0.5. The number of clusters, K, is assumed to be known beforehand in 3W-DPET, CE3, NEO-K-Means, FCM, RKM, Kernel K-Means, and SC, and it is set as the true number of classes in the dataset. In this case, the comparison may not be fair.

In our experiments, the performance of M3W and comparison algorithms is measured by three widely used cluster indices, including normalized mutual information (NMI) [43],

adjusted rand index (ARI) [44], and pairwise $F1$ ($F1$) [45], which are widely used in previous studies [46], [47], [48]. The higher the scores of the three measures, the better the clustering results. NMI and $F1$ can yield a value between 0 and 1, where 0 and 1 illustrate the inappropriate and appropriate clustering, respectively. The ARI value varies between -1 and 1. Random labeling may have an ARI near 0, and a perfect match has an ARI of 1.
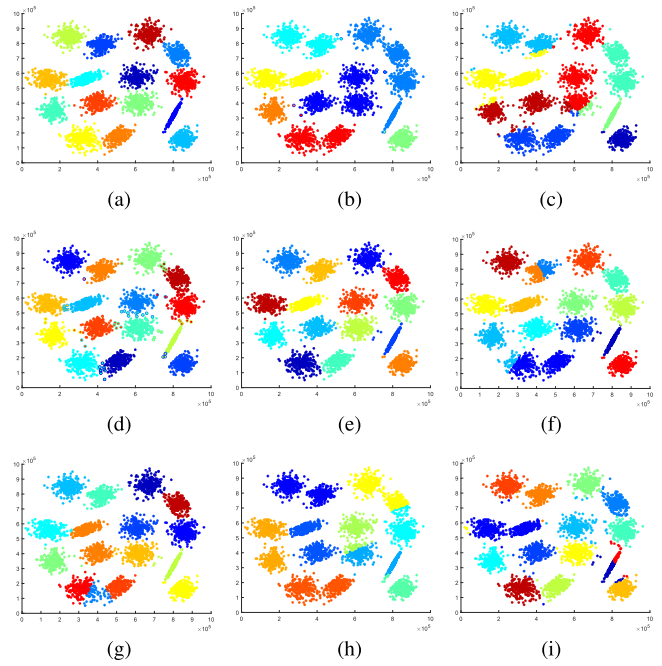
Fig. 7. Clustering results on T2. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.

Fig. 9. Clustering results on Ds2c2sc13. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.

Fig. 8. Clustering results on Pathbased. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.
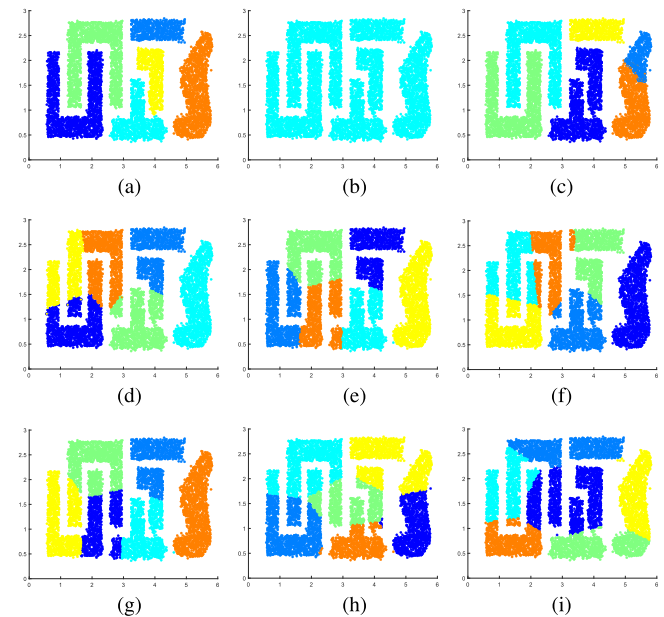
Fig. 10. Clustering results on T4. (a) M3W. (b) 3W-DBSCAN. (c) 3W-DPET. (d) CE3. (e) NEO-K-Means. (f) FCM. (g) RKM. (h) KnK-Means. (i) SC.
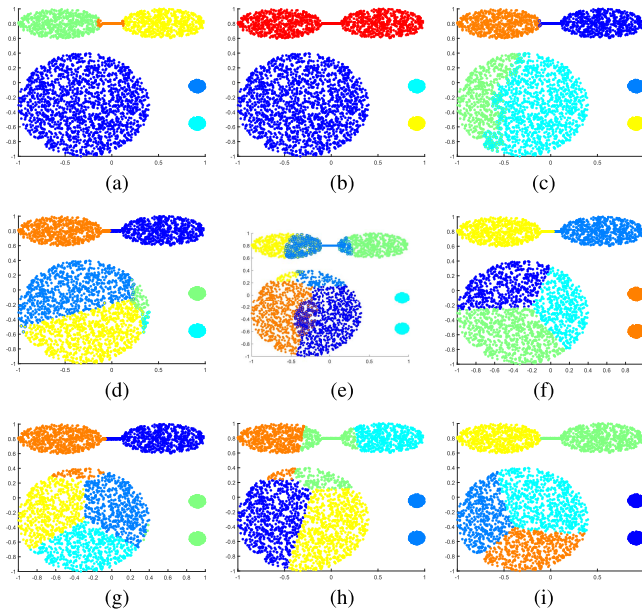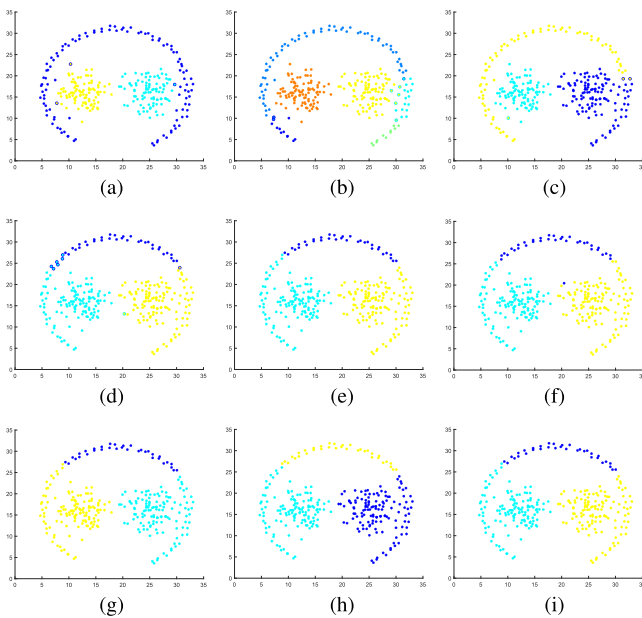
To more objectively reflect the performance of various algorithms, we run each algorithm with parameter ranges and take the best clustering result for each dataset. For CE3, NEO-K-Means, FCM, RKM, Kernel K-Means, and SC, we repeat each experiment with the same parameter input 25 times to reduce the influence of random centroid initialization and report the best result (from the best parameter input) in terms of the averages and standard deviations. Unlike them, 3W-DBSCAN, 3W-DPET, and our algorithm are deterministic algorithms. In other words, if one gives a specific parameter input, the clustering result will always be the same.

Therefore, the results do not have their standard deviations. The experiments are performed using a PC machine equipped with Intel(R) Core(TM)-i7-9700F CPU and 32-GB RAM in the MATLAB environment.

B. Experimental Results on Synthetic Datasets

In this section, we demonstrate the clustering results on eight synthetic datasets.

Triangle1 and Triangle2 have four Gaussian-distributed clusters with different variances. Figs. 3 and 4 demonstrate

TABLE III

PERFORMANCE COMPARISON OF M3W ON SYNTHETIC DATASETS

Algorithm	NMI	ARI	F1	Params	NMI	ARI	F1	Params
	Triangle1				Triangle2			
M3W	**1.000**	**1.000**	**1.000**	k=21, L=3	**0.987**	**0.990**	**0.993**	k=25, L=4
3W-DBSCAN	**1.000**	**1.000**	**1.000**	k=5, η=0.1, ε=0.1	0.974	0.986	**0.993**	k=15, η=0.4, ε=0.1
3W-DPET	**1.000**	**1.000**	**1.000**	K=4, k=5	0.968	0.981	**0.993**	K=4, k=5
CE3	0.948±0.000	0.970±0.000	**1.000±0.000**	K=4, q=6	0.859±0.000	0.917±0.000	**0.993±0.000**	K=4, q=14
NEO-K-Means	0.952±0.000	0.963±0.000	**1.000±0.000**	K=4, α=2, β=6	0.876±0.000	0.893±0.000	**0.993±0.000**	K=4, α=1.5, β=3
FCM	0.949±0.000	0.961±0.000	0.971±0.000	K=4, m=2	0.871±0.000	0.885±0.000	0.919±0.000	K=4, m=2
RKM	0.952±0.000	0.963±0.000	0.972±0.000	K=4, w_u=0.2, ϵ=1	0.869±0.000	0.882±0.000	0.916±0.000	K=4, w_u=0.4, ϵ=1
KnK-Means	0.978±0.108	0.984±0.185	0.988±0.118	K=4	0.886±0.056	0.901±0.065	0.938±0.004	K=4
SC	0.995±0.000	0.997±0.000	0.996±0.000	K=4, σ=0.5	0.901±0.001	0.919±0.001	0.953±0.001	K=4, σ=0.5
	S1				S2			
M3W	**0.989**	**0.989**	**0.990**	k=25, L=12	**0.946**	**0.938**	0.922	k=30, L=12
3W-DBSCAN	0.821	0.507	**0.990**	k=17, η=0.2, ε=0.1	0.000	0.000	0.922	k=20, η=1, ε=1
3W-DPET	0.819	0.587	**0.990**	K=15, k=20	0.656	0.337	0.922	K=15, k=20
CE3	0.979±0.025	0.982±0.076	**0.990±0.033**	K=15, q=8	0.898±0.019	0.840±0.057	0.854±0.068	K=15, q=8
NEO-K-Means	0.986±0.008	0.986±0.010	**0.990±0.046**	K=15, α=1, β=6	**0.946±0.008**	0.937±0.053	**0.941±0.028**	K=15, α=1, β=6
FCM	0.951±0.017	0.890±0.042	0.898±0.036	K=15, m=2	0.946±0.010	0.938±0.014	**0.942±0.032**	K=15, m=2.5
RKM	0.962±0.017	0.908±0.045	0.914±0.007	K=15, w_u=0.3, ϵ=1	0.922±0.024	0.901±0.055	0.908±0.049	K=15, w_u=0.4, ϵ=0.8
KnK-Means	0.861±0.017	0.662±0.029	0.691±0.089	K=15	0.820±0.034	0.622±0.072	0.654±0.082	K=15
SC	0.950±0.007	0.898±0.027	0.905±0.037	K=15, σ=2	0.905±0.021	0.845±0.062	0.856±0.050	K=15, σ=1.5
	T2				Pathbased			
M3W	**0.986**	**0.995**	**0.997**	k=24, L=2	**0.935**	**0.960**	**0.973**	k=9, L=4
3W-DBSCAN	0.873	0.809	**0.997**	k=20, η=0.9, ε=0.1	0.810	0.801	**0.973**	k=5, η=0.4, ε=0.1
3W-DPET	0.890	0.767	**0.997**	K=6, k=10	0.781	0.732	**0.973**	K=3, k=8
CE3	0.806±0.036	0.640±0.082	0.717±0.055	K=6, q=9	0.550±0.008	0.473±0.002	0.668±0.001	K=3, q=18
NEO-K-Means	0.789±0.028	0.614±0.028	0.706±0.011	K=6, α=1.5, β=6	0.549±0.001	0.464±0.001	0.660±0.000	K=3, α=0, β=6
FCM	0.758±0.017	0.515±0.051	0.615±0.032	K=6, m=3.5	0.535±0.000	0.460±0.000	0.657±0.000	K=3, m=2
RKM	0.727±0.013	0.480±0.042	0.588±0.007	K=6, w_u=0.2, ϵ=0.8	0.549±0.000	0.464±0.000	0.660±0.000	K=3, w_u=0.4, ϵ=1
KnK-Means	0.719±0.031	0.522±0.048	0.622±0.033	K=6	0.552±0.001	0.468±0.001	0.664±0.001	K=3
SC	0.778±0.042	0.522±0.069	0.622±0.032	K=6, σ=2	0.501±0.000	0.459±0.000	0.668±0.031	K=3, σ=0.5
	Ds2c2sc13				T4			
M3W	**0.994**	**0.994**	**0.995**	k=8, L=3	**1.000**	**1.000**	**1.000**	k=20, L=2
3W-DBSCAN	0.936	0.848	**0.995**	k=10, η=0.1, ε=0.1	0.000	0.000	0.322	k=20, η=1, ε=1
3W-DPET	0.948	0.842	**0.995**	K=13, k=13	0.924	0.872	0.897	K=6, k=7
CE3	0.778±0.019	0.567±0.033	0.605±0.010	K=13, q=16	0.713±0.004	0.615±0.019	0.685±0.000	K=6, q=17
NEO-K-Means	0.795±0.017	0.577±0.025	0.615±0.008	K=13, α=0.5, β=3	0.707±0.004	0.607±0.019	0.678±0.014	K=6, α=0, β=6
FCM	0.782±0.016	0.579±0.050	0.616±0.072	K=13, m=2	0.727±0.055	0.636±0.085	0.702±0.089	K=6, m=4
RKM	0.784±0.011	0.472±0.030	0.524±0.043	K=13, w_u=0.3, ϵ=1	0.707±0.005	0.607±0.021	0.678±0.003	K=6, w_u=0.1, ϵ=1
KnK-Means	0.792±0.037	0.626±0.081	0.586±0.060	K=13	0.629±0.045	0.511±0.063	0.600±0.037	K=6
SC	0.799±0.024	0.567±0.070	0.607±0.065	K=13, σ=1	0.588±0.024	0.393±0.020	0.504±0.046	K=6, σ=1.5

the clustering results of different clustering on two datasets. In Triangle2, the margins between adjacent clusters are smaller than those in Triangle1. Fig. 3 shows that M3W, 3W-DBSCAN, and 3W-DPET are able to distinguish perfectly clusters in Triangle1. Other algorithms all detect the general shape of each cluster, but there are some erroneously clustered points. Fig. 4 shows that the results of M3W, 3W-DBSCAN, and 3W-DPET are better than those of other algorithms. Note that the edges of some points are marked with one color, while their "faces" are marked with another. This means that these points belong to both of the color-coded clusters.

In the S1 and S2 datasets, there are 5000 points grouped around 15 clusters with varying levels of overlap. S2 has a higher level of overlap than S1. This means S1 is well clustered compared with S2. Figs. 5 and 6 show that M3W, CE3, NEO-K-Means, and SC get better clustering performance than other algorithms. It is noteworthy that 3W-DBSCAN yields an extremely bad result, where different clusters on S2 are merged into a single one. This is probably caused by the properties of DBSCAN. In DBSCAN, a density-based cluster is defined as a contiguous region with high density. DBSCAN (or 3W-DBSCAN) may create incorrect links between adjacent clusters when there is a very small margin between them. The erosion strategy employed by M3W along with the multistep association strategy solves this problem perfectly, as shown in Fig. 6.

A comparison of clustering results on T2 is shown in Fig. 7. In this dataset, a line-shaped cluster is located between two oval-shaped clusters. Our algorithm is the only one that can distinguish the general shapes of clusters. Also, a few points in the boundary regions are assigned to multiple clusters.

As shown in Fig. 8, Pathbased is challenging for most clustering algorithms. In this dataset, two Gaussian distributed clusters are surrounded by a half-circle cluster, and adjacent clusters are closely spaced (especially, the boundary between the half-circle cluster and the right Gaussian distributed cluster is vague and unclear). Although the counterparts fail to find the optimal structure of three clusters, our algorithm gives a much more satisfactory result.

TABLE IV
PERFORMANCE COMPARISON OF M3W ON REAL-WORLD DATASETS

Algorithm	NMI	ARI	F1	Params	NMI	ARI	F1	Params
	Glass				Dermatology			
M3W	**0.811**	**0.709**	**0.772**	$k=5, L=2$	**0.915**	0.849	0.882	$k=8, L=6$
3W-DBSCAN	0.665	0.458	0.616	$k=7, \eta=0.2, \varepsilon=0.1$	0.763	0.622	0.712	$k=15, \eta=0.5, \varepsilon=0.5$
3W-DPET	0.596	0.398	0.590	$K=6, k=7$	0.910	**0.892**	**0.917**	$K=6, k=6$
CE3	0.699±0.010	0.486±0.010	0.590±0.023	$K=6, q=18$	0.821±0.093	0.740±0.051	0.789±0.086	$K=6, q=15$
NEO-K-Means	0.720±0.010	0.531±0.015	0.625±0.017	$K=6, \alpha=-0.5, \beta=6$	0.877±0.038	0.738±0.018	0.791±0.128	$K=6, \alpha=0.5, \beta=6$
FCM	0.762±0.000	0.570±0.000	0.656±0.000	$K=6, m=2$	0.784±0.000	0.657±0.000	0.729±0.020	$K=6, m=2$
RKM	0.730±0.010	0.537±0.014	0.630±0.009	$K=6, w_u=0.1, \epsilon=0.8$	0.872±0.034	0.798±0.009	0.840±0.165	$K=6, w_u=0.2, \epsilon=0.8$
KnK-Means	0.645±0.062	0.460±0.121	0.581±0.026	$K=6$	0.856±0.047	0.821±0.026	0.856±0.096	$K=6$
SC	0.726±0.033	0.554±0.043	0.646±0.034	$K=6, \sigma=1.5$	0.868±0.017	0.705±0.011	0.763±0.056	$K=6, \sigma=0.5$
	Digits				MSRA			
M3W	**0.879**	**0.787**	0.718	$k=14, L=11$	**0.836**	**0.622**	**0.705**	$k=22, L=8$
3W-DBSCAN	0.820	0.710	**0.744**	$k=15, \eta=0.6, \varepsilon=0.5$	0.692	0.517	0.644	$k=16, \eta=0.2, \varepsilon=0.8$
3W-DPET	0.813	0.543	0.610	$K=10, k=15$	0.310	0.048	0.553	$K=12, k=20$
CE3	0.749±0.029	0.652±0.061	0.689±0.057	$K=10, q=6$	0.546±0.053	0.276±0.078	0.196	$K=12, q=8$
NEO-K-Means	0.742±0.009	0.645±0.027	0.683±0.003	$K=10, \alpha=0.5, \beta=6$	0.582±0.031	0.314±0.051	0.352±0.038	$K=12, \alpha=0, \beta=6$
FCM	0.405±0.030	0.231±0.035	0.339±0.012	$K=10, m=2.5$	0.258±0.008	0.119±0.007	0.385±0.032	$K=12, m=2$
RKM	0.000±0.000	0.000±0.000	0.181±0.000	$K=10, w_u=0.4, \epsilon=0.8$	0.655±0.031	0.440±0.037	0.490±0.037	$K=12, w_u=0.1, \epsilon=0.8$
KnK-Means	0.750±0.030	0.653±0.062	0.690±0.022	$K=10$	0.604±0.020	0.351±0.031	0.411±0.045	$K=12$
SC	0.732±0.002	0.665±0.003	0.698±0.002	$K=10, \sigma=1$	0.614±0.025	0.408±0.036	0.460±0.015	$K=12, \sigma=1$
	Segmentation				Optdigits			
M3W	**0.728**	**0.411**	**0.533**	$k=30, L=6$	**0.871**	**0.814**	**0.833**	$k=20, L=12$
3W-DBSCAN	0.607	0.253	0.417	$k=18, \eta=0.1, \varepsilon=0.2$	0.845	0.788	0.810	$k=17, \eta=0.4, \varepsilon=0.6$
3W-DPET	0.512	0.284	0.428	$K=7, k=9$	0.817	0.695	0.733	$K=10, k=5$
CE3	0.557±0.101	0.393±0.118	0.494±0.038	$K=7, q=8$	0.747±0.034	0.680±0.078	0.713±0.041	$K=10, q=5$
NEO-K-Means	0.518±0.031	0.337±0.040	0.464±0.043	$K=7, \alpha=0.5, \beta=6$	0.745±0.007	0.674±0.026	0.707±0.022	$K=10, \alpha=-0.5, \beta=3$
FCM	0.516±0.035	0.391±0.039	0.479±0.034	$K=7, m=2$	0.406±0.017	0.224±0.032	0.339±0.032	$K=10, m=4$
RKM	0.561±0.050	0.401±0.061	0.498±0.040	$K=7, w_u=0.1, \epsilon=1$	0.000±0.000	0.000±0.000	0.182±0.000	$K=10, w_u=0.4, \epsilon=0.8$
KnK-Means	0.518±0.060	0.360±0.062	0.464±0.035	$K=7$	0.768±0.009	0.686±0.009	0.719±0.031	$K=10$
SC	0.390±0.027	0.274±0.047	0.384±0.037	$K=7, \sigma=2$	0.731±0.007	0.629±0.023	0.668±0.039	$K=10, \sigma=0.5$
	Statlog				Pendigits			
M3W	**0.693**	**0.686**	**0.752**	$k=25, L=12$	**0.865**	**0.809**	**0.827**	$k=30, L=8$
3W-DBSCAN	0.620	0.432	0.580	$k=20, \eta=0.7, \varepsilon=0.1$	0.732	0.646	0.683	$k=20, \eta=1, \varepsilon=0.1$
3W-DPET	0.556	0.318	0.510	$K=6, k=20$	0.706	0.514	0.578	$K=10, k=5$
CE3	0.584±0.047	0.516±0.066	0.611±0.067	$K=6, q=19$	0.692±0.008	0.612±0.039	0.654±0.036	$K=10, q=17$
NEO-K-Means	0.611±0.062	0.528±0.086	0.616±0.053	$K=6, \alpha=-0.5, \beta=6$	0.691±0.012	0.595±0.038	0.637±0.027	$K=10, \alpha=-1, \beta=6$
FCM	0.584±0.000	0.509±0.000	0.600±0.000	$K=6, m=2$	0.638±0.026	0.523±0.039	0.577±0.000	$K=10, m=2$
RKM	0.612±0.058	0.530±0.090	0.617±0.043	$K=6, w_u=0.4, \epsilon=1$	0.266±0.095	0.082±0.082	0.236±0.049	$K=10, w_u=0.1, \epsilon=0.8$
KnK-Means	0.617±0.048	0.530±0.069	0.617±0.071	$K=6$	0.690±0.012	0.581±0.049	0.624±0.023	$K=10$
SC	0.615±0.013	0.500±0.052	0.592±0.026	$K=6, \sigma=1$	0.679±0.014	0.574±0.037	0.618±0.023	$K=10, \sigma=1$
	Htru				Shuttle			
M3W	**0.554**	**0.709**	**0.951**	$k=25, L=12$	**0.809**	**0.773**	**0.849**	$k=35, L=5$
3W-DBSCAN	0.446	0.554	0.944	$k=15, \eta=0.3, \varepsilon=0.1$	0.015	0.001	0.558	$k=8, \eta=0.1, \varepsilon=0.4$
3W-DPET	0.436	0.519	0.941	$K=2, k=7$	0.449	0.383	0.656	$K=3, k=20$
CE3	0.353±0.000	0.530±0.000	0.907±0.000	$K=2, q=18$	0.010±0.003	0.001±0.000	0.558±0.014	$K=3, q=17$
NEO-K-Means	0.342±0.000	0.530±0.001	0.906±0.000	$K=2, \alpha=-0.5, \beta=6$	0.000±0.139	0.001±0.169	0.685±0.050	$K=3, \alpha=3, \beta=6$
FCM	0.269±0.000	0.343±0.000	0.819±0.000	$K=2, m=4$	0.236±0.001	0.244±0.001	0.526±0.000	$K=3, m=4$
RKM	0.365±0.000	0.579±0.000	0.929±0.000	$K=2, w_u=0.7, \epsilon=0.8$	0.384±0.102	0.364±0.099	0.591±0.061	$K=3, w_u=0.5, \epsilon=0.8$
KnK-Means	0.290±0.000	0.400±0.000	0.848±0.000	$K=2$	0.309±0.038	0.293±0.027	0.582±0.008	$K=3$
SC	0.376±0.000	0.571±0.000	0.917±0.000	$K=2, \sigma=0.5$	0.010±0.000	0.002±0.000	0.558±0.000	$K=3, \sigma=0.5$

Ds2c2sc13 consists of 13 clusters with varied sizes and shapes. As shown in Fig. 9, M3W outperforms other algorithms on the dataset. Only a few points in two small regions are assigned to multiple clusters.

T4 is composed of six complex-shaped clusters. Fig. 10 shows that only M3W can generate the optimal structure of clusters on T4.

As shown in Figs. 3–8, when nearby clusters have unclear and vague boundaries, the proposed algorithm still performs well. M3W is seen as a DBSCAN-like algorithm. For density-based clustering methods, a small margin between adjacent clusters may lead to the undesirable connectivity of the clusters. Unlike other density-based clustering algorithms, the proposed algorithm overcomes this problem effectively.

M3W uses the erosion strategy to ensure that the initial core regions of nearby clusters are clearly separated. Figs. 8–9 show that our algorithm is very effective in finding clusters with complex shapes. The main reason is that a multistep association strategy is adopted to ensure that more information is obtained gradually to improve the probability of being correctly assigned, thereby obtaining satisfactory results. To further explain the reason behind M3W's superiority, we use the Pathbased dataset for the example and present a more detailed experimental analysis in the Supplementary Material.

The quantitative results of these synthetic datasets are illustrated in Table III. It can be examined from the experimental results that the proposed algorithm clearly outperforms other
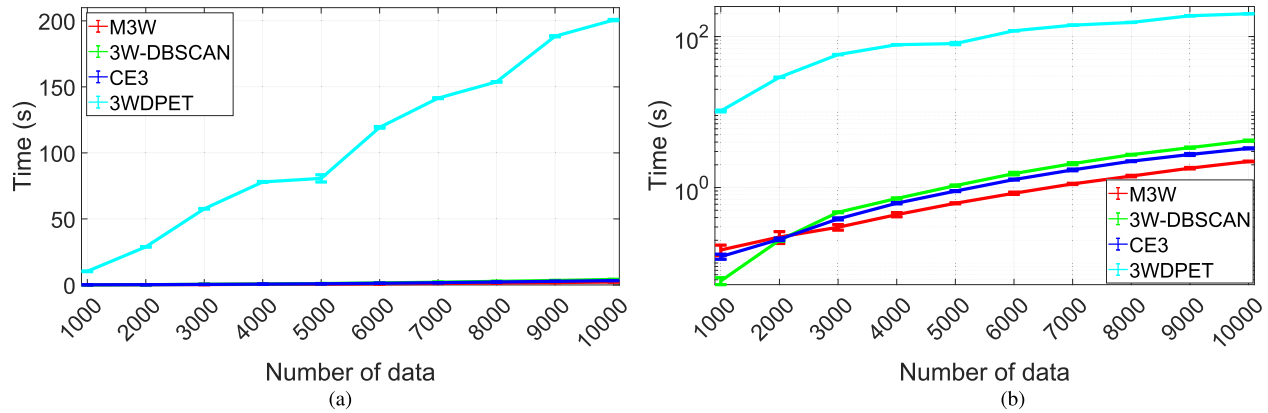
Fig. 11. Running time comparison. (a) Linear scale on the *y*-axis. (b) Log scale on the *y*-axis.

algorithms on these synthetic datasets. Additional results are presented in the Supplementary Material, including accuracy (ACC) and two internal clustering validation measures (S_Dbw index and local cores-based cluster validity (LCCV) index).

C. Experimental Results on Real-World Datasets

In Table IV, we list the clustering performance of the proposed algorithm and eight competitors on real-world datasets. The highest scores for each dataset are highlighted in bold type in Table IV. The NMI scores of M3W are higher than those of other algorithms in all test cases. In addition, the proposed algorithm achieves the best results in terms of ARI and $F1$ on at least eight real-world datasets. Specifically, on the Glass dataset, the result obtained by the proposed algorithm is significantly superior to the second-best result. On the Dermatology dataset, although the ARI and $F1$ values of M3W are slightly lower than those of 3W-DPET, it outperforms other algorithms in terms of NMI. The results on Digits show that our algorithm exceeds the second-best one by over 5% points in terms of NMI and $F1$. M3W and 3W-DBSCAN are extremely close in $F1$, with a difference of just 0.03. On the MSRA dataset, M3W performs better than other algorithms in terms of all evaluation metrics. The proposed algorithm shows an advantage on Segmentation. On the Optdigits dataset, the proposed technique is slightly superior to other competing algorithms. The proposed algorithm outperforms other competing algorithms on the Statlog dataset. As a result of the Pendigits dataset, our algorithm has over 10% higher performance than the second-best one in all evaluation metrics. On the Htru and Shuttle datasets, M3W is significantly superior to other algorithms. Additional results are presented in the Supplementary Material, including Accuracy (ACC) and two internal clustering validation measures (S_Dbw Index and LCCV index). In addition, we use t-SNE [49] to visualize the clustering results on real-world datasets in the Supplementary Material.

In summary, the superiority of M3W can be attributed to two factors: 1) the progressive erosion and 2) the multistep allocation strategy. The erosion process can reveal the natural structure of the latent clusters. The multistep

allocation fully takes advantage of the neighborhood information, thereby refining the clustering results. In contrast, other three-way clustering algorithms do not sufficiently consider the neighborhood information of the data and the cluster structure. This may affect the clustering performance.

D. Running Time

This section compares the running times of M3W and other three-way clustering algorithms (3W-DBSCAN, 3WDPET, and CE3)[7] on synthetic datasets with different numbers of data instances ($N = 1000:1000:10000$). For a fair comparison, the number of neighbors (a parameter common to all algorithms) is kept at 20. To determine the running time, we use the average and standard deviation of 25 repeated experiments. We perform all experiments in the MATLAB environment on a PC machine containing an Intel(R) Core(TM)-i7-9700F CPU and 32-GB RAM.

Fig. 11 presents the averages and standard deviations of the running times for M3W, 3W-DBSCAN, 3WDPET, and CE3. The error bars in Fig. 11 indicate the standard deviations. We see that the 3WDPET is significantly slower than other algorithms (3W-DBSCAN, 3WDPET, and M3W) in Fig. 11(a). To provide a further comparison, Fig. 11(b) shows this same result using a log scale on the *y*-axis. The following observations can be drawn from Fig. 11(b). M3W is faster in most cases than 3W-DBSCAN and CE3. As the number of data instances increases, the advantage that M3W shows is more pronounced, indicating that it is more scalable than other three-way clustering algorithms.

VI. CONCLUSION

This article presents a new three-way clustering algorithm, called M3W. A progressive erosion technique using dynamic density estimation is adopted to generate a multilevel structure of the data. We define a multistep allocation strategy by

---

[7]To make a fair comparison, we chose other three-way clustering algorithms as our comparison partners in the running time experiments. A three-way clustering algorithm is often slower than a two-way clustering algorithm, because it takes more time to compute and store cluster representations with interval sets (see Section I-A). In the Supplementary Material, we provide a comparison of all clustering algorithms in terms of running time.

integrating the idea of sequential three-way decisions into the clustering process. With the aid of the multilevel structure of the data, the allocation strategy may exploit more available information, leading to better clustering. Experiments on both synthetic and real datasets demonstrate that M3W has satisfactory performance.

In future work, we plan to extend M3W to online learning. Determining the parameters (i.e., k and L) automatically is also an interesting and worth-exploring issue. Moreover, we plan on developing a general framework for sequential three-way clustering.

REFERENCES

[1] H. Liu, E. Li, X. Liu, K. Su, and S. Zhang, "Anomaly detection with kernel preserving embedding," *ACM Trans. Knowl. Discovery From Data*, vol. 15, no. 5, pp. 1–18, Oct. 2021.

[2] W. Ding, J. Nayak, B. Naik, D. Pelusi, and M. Mishra, "Fuzzy and real-coded chemical reaction optimization for intrusion detection in industrial big data environment," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4298–4307, Jun. 2021.

[3] W. Ding *et al.*, "An unsupervised fuzzy clustering approach for early screening of COVID-19 from radiological images," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 2902–2914, Aug. 2022.

[4] J. Liu, X. Liu, Y. Yang, X. Guo, M. Kloft, and L. He, "Multiview subspace clustering via co-training robust data representation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 9, 2021, doi: 10.1109/TNNLS.2021.3069424.

[5] Q. Feng, L. Chen, C. P. Chen, and L. Guo, "Deep fuzzy clustering—A representation learning approach," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 7, pp. 1420–1433, Jul. 2020.

[6] H. Yu and Y. Wang, "Three-way decisions method for overlapping clustering," in *Proc. 8th Int. Conf. Rough Sets Current Trends Comput.* Chengdu, China, 2012, pp. 277–286.

[7] Y. Yao, "Three-way decisions with probabilistic rough sets," *Inf. Sci.*, vol. 180, no. 3, pp. 341–353, 2010.

[8] Y. Yao, "Tri-level thinking: Models of three-way decision," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 5, pp. 947–959, May 2020.

[9] Y. Yao and X. Deng, "Sequential three-way decisions with probabilistic rough sets," in *Proc. IEEE 10th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI-CC)*, Aug. 2011, pp. 120–125.

[10] Y. Yao, "Granular computing and sequential three-way decisions," in *Proc. 8th Int. Conf. Rough Sets Knowl. Technol.* Halifax, NS, Canada, 2013, pp. 16–27.

[11] Y. Yao and C. Gao, "Statistical interpretations of three-way decisions," in *Proc. 10th Int. Conf. Rough Sets Knowl. Technol.* Tianjin, China, 2015, pp. 309–320.

[12] X. Wang, P. Wang, X. Yang, and Y. Yao, "Attribution reduction based on sequential three-way search of granularity," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 5, pp. 1439–1458, May 2021.

[13] G. Lang, D. Miao, and H. Fujita, "Three-way group conflict analysis based on Pythagorean fuzzy set theory," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 3, pp. 447–461, Mar. 2020.

[14] X. Yue, J. Zhou, Y. Yao, and D. Miao, "Shadowed neighborhoods based on fuzzy rough transformation for three-way classification," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 978–991, May 2020.

[15] P. Wang, H. Shi, X. Yang, and J. Mi, "Three-way k-means: Integrating k-means and three-way decision," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 10, pp. 2767–2777, Oct. 2019.

[16] H. Averbuch-Elor, N. Bar, and D. Cohen-Or, "Border-peeling clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 7, pp. 1791–1797, Jul. 2020.

[17] M. Du, R. Wang, R. Ji, X. Wang, and Y. Dong, "ROBP a robust border-peeling clustering using Cauchy kernel," *Inf. Sci.*, vol. 571, pp. 375–400, Sep. 2021.

[18] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.

[19] P. Lingras and C. West, "Interval set clustering of web users with rough K-means," *J. Intell. Inf. Syst.*, vol. 23, no. 1, pp. 5–16, Jul. 2004.

[20] H. Yu and Q. Zhou, "A cluster ensemble framework based on three-way decisions," in *Proc. 8th Int. Conf. Rough Sets Knowl. Technol.* Halifax, NS, Canada, 2013, pp. 302–312.

[21] H. Yu and G. Wang, "An efficient gradual three-way decision cluster ensemble approach," in *Proc. 17th Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst.* Cádiz, Spain, 2018, pp. 711–723.

[22] P. Wang and X. Yang, "Three-way clustering method based on stability theory," *IEEE Access*, vol. 9, pp. 33944–33953, 2021.

[23] H. Yu, C. Zhang, and G. Wang, "A tree-based incremental overlapping clustering method using the three-way decision theory," *Knowl.-Based Syst.*, vol. 91, pp. 189–203, Jan. 2016.

[24] H. Yu, P. Jiao, Y. Yao, and G. Wang, "Detecting and refining overlapping regions in complex networks with three-way decisions," *Inf. Sci.*, vol. 373, pp. 21–41, Dec. 2016.

[25] H. Yu, "A framework of three-way cluster analysis," in *Proc. Int. Joint Conf. Rough Sets* Olsztyn, Poland, 2017, pp. 300–312.

[26] H. Yu, X. Wang, G. Wang, and X. Zeng, "An active three-way clustering method via low-rank matrices for multi-view data," *Inf. Sci.*, vol. 507, pp. 823–839, Jan. 2020.

[27] C. Zhu, L. Ma, P. Wang, and D. Miao, "Multi-view and multi-label method with three-way decision-based clustering," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.* Nanjing, China, 2020, pp. 69–80.

[28] P. Wang and Y. Yao, "CE3: A three-way clustering method based on mathematical morphology," *Knowl.-Based Syst.*, vol. 155, pp. 54–65, Sep. 2018.

[29] H. Yu, L. Chen, J. Yao, and X. Wang, "A three-way clustering method based on an improved DBSCAN algorithm," *Phys. A, Stat. Mech. Appl.*, vol. 535, Dec. 2019, Art. no. 122289.

[30] K. Zhang, "A three-way c-means algorithm," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105536.

[31] H. Yu, L. Chen, and J. Yao, "A three-way density peak clustering method based on evidence theory," *Knowl.-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106532.

[32] M. K. Afridi, N. Azam, and J. Yao, "Variance based three-way clustering approaches for handling overlapping clustering," *Int. J. Approx. Reasoning*, vol. 118, pp. 47–63, Mar. 2020.

[33] H. Yu, Z. Chang, G. Wang, and X. Chen, "An efficient three-way clustering algorithm based on gravitational search," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 5, pp. 1003–1016, May 2020.

[34] X. Jia, Y. Rao, W. Li, S. Yang, and H. Yu, "An automatic three-way clustering method based on sample similarity," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 5, pp. 1545–1556, May 2021.

[35] L. Breiman, W. Meisel, and E. Purcell, "Variable kernel estimates of multivariate densities," *Technometrics*, vol. 19, no. 2, pp. 135–144, 1977.

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

[37] M. Du, S. Ding, and Y. Xue, "A robust density peaks clustering algorithm using fuzzy neighborhood," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 7, pp. 1131–1140, 2018.

[38] M. Ester *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining* Portland, OR, USA, 1996, pp. 226–231.

[39] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Proc. 17th Pacific–Asia Conf. Knowl. Discovery Data Mining*, 2013, pp. 160–172.

[40] J. J. Whang, Y. Y. Hou, D. F. Gleich, and I. S. Dhillon, "Non-exhaustive, overlapping clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2644–2659, Nov. 2019.

[41] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[42] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[43] M. Meilă, "Comparing clusterings—An information based distance," *J. Multivariate Anal.*, vol. 98, no. 5, pp. 873–895, May 2007.

[44] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[45] D. Pfitzner, R. Leibbrandt, and D. Powers, "Characterization and evaluation of similarity measures for pairs of clusterings," *Knowl. Inf. Syst.*, vol. 19, no. 3, pp. 361–394, Jun. 2009.

[46] M. Lu, X.-J. Zhao, L. Zhang, and F.-Z. Li, "Semi-supervised concept factorization for document clustering," *Inf. Sci.*, vol. 331, pp. 86–98, Feb. 2016.

[47] F. Nie, X. Dong, L. Tian, R. Wang, and X. Li, "Unsupervised feature selection with constrained $\ell_{2,0}$-norm and optimized graph," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1702–1713, Apr. 2022.

[48] Y. Lu, Y.-M. Cheung, and Y. Y. Tang, "Self-adaptive multiprototype-based competitive learning approach: A k-means-type algorithm for imbalanced data clustering," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1598–1612, Mar. 2021.

[49] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

Jiarui Sun received the bachelor's degree from Jiangsu Normal University, Xuzhou, China, in 2019, where he is currently pursuing the master's degree with the School of Computer Science and Technology.

His research interests include big data analysis and data stream clustering.

Mingjing Du (Member, IEEE) received the Ph.D. degree in computer science from the China University of Mining and Technology, Beijing, China, in 2018.

He is currently an Associate Professor with the School of Computer Science and Technology, Jiangsu Normal University, Xuzhou, China. His research results have expounded in over ten publications at peer-reviewed journals. His current research interests include cluster analysis and three-way decisions. For more information, see https://dumingjing.github.io/

Jingqi Zhao received the B.A. degree from Jiangsu Normal University, Xuzhou, China, in 2012, where she is currently pursuing the master's degree with the School of Fine Arts.

Her research interests include 3-D reconstruction and cluster analysis.

Yongquan Dong received the Ph.D. degree in computer science from Shandong University, Jinan, China, in 2010.

He is currently a Professor with the School of Computer Science and Technology, Jiangsu Normal University, Xuzhou, China. His research interests include web information integration and web data management.