Article

An Improved Three-Way Clustering Based on Ensemble Strategy

Tingfeng Wu [1], Jiachen Fan [1] and Pingxin Wang [2],*

[1] School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212003, China;
    202070018@stu.just.edu.cn (T.W.); 202070026@stu.just.edu.cn (J.F.)
[2] School of Science, Jiangsu University of Science and Technology, Zhenjiang 212003, China
*   Correspondence: wangpingxin@just.edu.cn

Abstract: As a powerful data analysis technique, clustering plays an important role in data mining. Traditional hard clustering uses one set with a crisp boundary to represent a cluster, which cannot solve the problem of inaccurate decision-making caused by inaccurate information or insufficient data. In order to solve this problem, three-way clustering was presented to show the uncertainty information in the dataset by adding the concept of fringe region. In this paper, we present an improved three-way clustering algorithm based on an ensemble strategy. Different to the existing clustering ensemble methods by using various clustering algorithms to produce the base clustering results, the proposed algorithm randomly extracts a feature subset of samples and uses the traditional clustering algorithm to obtain the diverse base clustering results. Based on the base clustering results, labels matching is used to align all clustering results in a given order and voting method is used to obtain the core region and the fringe region of the three way clustering. The proposed algorithm can be applied on the top of any existing hard clustering algorithm to generate the base clustering results. As examples for demonstration, we apply the proposed algorithm on the top of K-means and spectral clustering, respectively. The experimental results show that the proposed algorithm is effective in revealing cluster structures.

Keywords: ensemble clustering; three-way decision; three-way clustering; voting

MSC: 68T37

1. Introduction

Clustering aims to classify similar elements into the same cluster and dissimilar elements into different clusters by calculating the certain similarity between all elements, where the elements in the same cluster must have a high similarity and the elements in the different clusters have a low similarity [1,2]. As a key technology of machine learning, clustering is widely used in different fields such as information granulation [3–5], information fusion [6–8], attribute reduction [9–12], feature selection [13–15], etc. Many clustering methods have been developed over the past decades. Most of the existing clustering algorithms can be divided into five categories: the partition-based method [16,17], hierarchy-based method [18–20], density-based method [21–24], grid-based method [25] and model-based method [26–28].

Although there are many different clustering algorithms, the lack of prior knowledge means that clustering analysis remains a very challenging problem. It has been accepted that a single clustering approach cannot always describe the structural characteristics accurately, even if the same clustering approach cannot obtain good clusters because of the different initial parameters as well. To avoid this problem, ensemble clustering [29–31] has been developed to improve the robustness, stability, and quality of a clustering solution. Compared with a single clustering approach, the clustering results obtained by ensemble clustering approaches have better stability, robustness and accuracy.

In recent years, ensemble clustering has received more attention and many new ensemble clustering approaches have been developed [32–35].

The above works have obtained good performance in solving a clustering ensemble problem. However, most of the existing ensemble clustering results are hard clustering, in which an element belongs to one cluster or does not belong to one cluster, and there is a clear boundary between different clusters. Hard clustering algorithms often bring a higher decision risk when the sample's information is insufficient. In order to solve this problem, three-way decision [36] was proposed to describe the uncertainty of the information. The main idea of three-way decision is to divide a universe into three disjoint regions and make different strategies for different regions [37–39]. Many soft computing models for learning uncertain concepts, such as rough sets [40], fuzzy sets [41], shadowed sets [42] and concept learning [43,44] can be reinvestigated within the framework of three-way decision. By integrating three-way decision with clustering, Yu [45–47] proposed three-way clustering, which uses a core region and the fringe region to represent a cluster. These two sets divide the universe into three parts, where there are three types of relationships between the objects and cluster, namely objects belonging to the cluster, objects not belonging to the cluster, and objects partially belonging to the cluster. Recently, three-way clustering has attracted much research, and many three-way clustering algorithms have been developed [48–56].

This paper aims at presenting an improved three-way clustering algorithm based on an ensemble strategy to solve the problem of inaccurate decision-making caused by inaccurate information or insufficient data. Different to the existing clustering ensemble methods by using various clustering algorithms to produce the base clustering results, the proposed algorithm randomly extracts a feature subset of samples and uses the traditional clustering algorithm to obtain the diverse base clustering results. Based on the base clustering results, we develop a three-way clustering method by using the voting method. The main process of the proposed algorithm has two steps. Firstly, we use parts of the sample's features to obtain the base clustering results. Secondly, we use label matching to align all clustering results to a given order and the voting method to obtain the core region and the fringe region of three-way clustering. The sample is assigned into the core region of a corresponding cluster when the frequency of the sample in the same cluster is more than the given threshold. The difference between the union of the cluster with the same labels and the core region are regarded as the fringe region of the specific cluster. Therefore, a three-way clustering is naturally formed. The proposed strategy can be applied on the top of any existing hard clustering algorithm. Three-way ensemble K-means and NJW are given as examples for demonstration in this paper.

The remainder of this paper is organized as follows. In Section 2, we mainly introduce the concepts of ensemble clustering and three-way clustering. The process of the proposed algorithm is presented in Section 3. The performances of the proposed three-way ensemble clustering algorithm are illustrated through some UCI datasets in Section 4. Conclusions and future works are given in Section 5.

2. Related Work

In this section, we review some concepts and related works about ensemble clustering and three-way clustering.

2.1. Ensemble Clustering

Each clustering algorithm has its unique method for discovering the data structure. Different clustering algorithms can obtain different clustering results even if the data are the same. A single clustering algorithm cannot deal with all types of data structure. It is also difficult to choose a specialized clustering algorithm because of the insufficient prior class information. Hence, researchers are devoted to integrating multiple clustering results to one clustering result, which is called ensemble clustering. Compared with a single

clustering algorithm, ensemble clustering algorithm can obtain a better clustering result with a higher performance of robustness, stability and quality.

The concept of ensemble clustering was first proposed by Strehl and Ghosh [29], who combined cluster labels without accessing the original features. Wang et al. [57] developed an ensemble clustering method for probabilities accumulation by considering factors such as cluster size, sample dimension and density. Punera and Ghosh [58] proposed several consensus algorithms that were suitable for soft clustering by extending the relatively hard clustering approaches. Sevillano et al. [59] presented a set of fuzzy consensus functions which can combine multiple soft clustering results into a final soft clustering result, using the application of positional and confidence voting techniques. Li et al. [60] developed an ensemble clustering algorithm based on a sample's stability. Yu et al. [55] presented a framework of three-way ensemble clustering based on Spark and proposed a consensus clustering algorithm based on cluster units.

In general, ensemble clustering can be roughly divided into two stages: base cluster generation and base cluster aggregation. Base cluster generation is the first step of the ensemble clustering algorithm, and a set of clusters which will be combined should be generated. There are no restrictions on how to achieve base clusters, so we may have many approaches, such as using different clustering algorithms or using the same clustering algorithm with different parameters. In this paper, we mainly research the process of base cluster aggregation and how to convert hard clustering results into a soft clustering result. The process of ensemble clustering is shown in Figure 1.
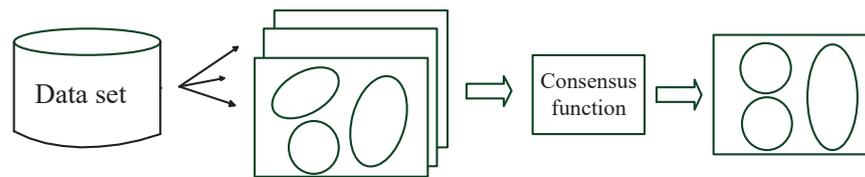

Figure 1. Flow chart of ensemble clustering.

2.2. Three-Way Clustering

Three-way decision [36] is an extension of two-way decision, in which a definite decision is given to the elements with definite information and a deferment decision is adopted when the elements' information is insufficient to avoid decision risk. Three-way decision uses three disjointed regions to represent a set, namely the positive region, the negative region and the boundary region, which correspond to the acceptance decision, the rejection decision and the delayed decision. Inspired by the idea of three-way decision, Yu [45] presented the framework of three-way clustering by combining the clustering approach and three-way decision.

We introduce some basic knowledge about three-way clustering. Given a set $U = \{x_1, x_2, \cdots, x_n\}$, which has n elements, the clustering result can be denoted as $\mathbb{C} = \{C^1, C^2, \cdots, C^K\}$ by using the hard clustering algorithm. In contrast to the hard clustering representation, the three-way cluster C^i is represented as a pair of sets:

$$C^i = \{Co(C^i), Fr(C^i)\}, \tag{1}$$

where $Co(C^i)$ is the core region of cluster C^i and $Fr(C^i)$ is the fringe region of cluster C^i. The third region is trivial region $Tr(C^i) = U - Co(C^i) - Fr(C^i)$, which can be expressed as the complement of the union of $Co(C^i)$ and $Fr(C^i)$.

We summarize some concepts of three-way clustering. If the element $v \in Co(C^i)$, v must belong to the cluster C^i; if the element $v \in Fr(C^i)$, v might belong to the cluster C^i; if

the element $v \in Tr(C^i)$, v does not belong to the cluster C^i. The above three subsets obey the following properties:

$$Co(C^i) \cap Fr(C^i) = \varnothing, \tag{2}$$
$$Co(C^i) \cap Tr(C^i) = \varnothing, \tag{3}$$
$$Fr(C^i) \cap Tr(C^i) = \varnothing, \tag{4}$$
$$Tr(C^i) \cup Co(C^i) \cup Fr(C^i) = U. \tag{5}$$

When $Fr(C^i) = \varnothing$, it is obvious to find that the cluster C^i can be only represented by the core region $Co(C^i)$, and it is a hard clustering result. So the hard clustering result is a special case of a three-way clustering result in a certain situation.

In this paper, we adopt the following three conditions:

(1) $Co(C^i) \neq \varnothing, i = 1, 2, \cdots, k$,

(2) $\bigcup\limits_{i=1}^{K} (Co(C^i) \cup Fr(C^i)) = U$,

(3) $Co(C^i) \cap Co(C^j) = \varnothing, i \neq j$.

Condition (1) means that each cluster must have elements. Condition (2) indicates that an element v belongs to more than one cluster. Condition (3) demands that the core region of each cluster must be disjointed. Therefore, we can represent the three-way clustering results as:

$$\mathbb{C} = \{(Co(C^i), Fr(C^i)), (Co(C^i), Fr(C^i)), \cdots, (Co(C^K), Fr(C^K))\}.$$

Some three-way clustering algorithms were developed since three-way clustering was proposed. Wang [49] proposed a three-way clustering framework by combining ideas of three-way decision and erosion and dilation from mathematical morphology. Wang et al. [50] improved the K-means algorithm and then developed a three-way k-means method. Zhang [61] presented a three-way c-means clustering algorithm by integrating the three-way weight and three-way assignment. Yu et al. [47] proposed an efficient three-way clustering algorithm based on the idea of universal gravitation, which can adjust the thresholds automatically in the process of clustering. Jia et al. [62] developed an automatic three-way clustering approach by combining the proposed threshold selection based on the roughness degree using sample's similarity and the cluster number selection method.

3. The Proposed Three-Way Clustering Based on Ensemble Strategy

Three-way clustering was presented to show the uncertainty information in the dataset by adding the concept of a fringe region. Although many three-way clustering algorithms have obtained good performances, there is still much room for improving the methods. We present an improved three-way clustering algorithm based on an ensemble strategy (TWCE for short) in this section. Compared with the existing algorithms, the proposed algorithm randomly extracts a feature subset of samples and uses the traditional clustering algorithm to obtain the diverse base clustering results. The computational complexity is lower than that of the existing clustering ensemble methods by using all the sample's features. In addition, the proposed strategy can be applied on the top of any existing hard clustering algorithm. The process of the proposed algorithm has three steps: generation of base clustering, labels matching and results of three-way clustering.

3.1. Generation of Base Clustering Results

The first task in the clustering ensemble is to obtain a set of base clustering results. There are many approaches to generating base clustering clusters, among which utilizing different clustering algorithms is the most commonly used strategy. Each clustering algorithm has its own specific view on how to discover the underlying structure of a dataset. Therefore, multiple clustering algorithms can be used to generate diverse base clustering results. Another commonly used method to obtain the base clustering results is to use one

clustering algorithm with different parameters. For example, base clustering results can be obtained by setting different numbers of clusters and the initial centers of the K-means type of algorithms. The above methods use all the sample's features during the process of clustering, which will take a lot of time for a multi-dimensional dataset.

Different from the existing clustering ensemble method, the proposed TWCE algorithm uses parts of the sample's features to obtain the base clustering results. For a multi-dimensional dataset, different subsets of features try to describe the dataset from different views. Thus, a set of diverse clustering results will be obtained when distinguishing subsets of features are utilized. Suppose U is a dataset with m features, we randomly extract parts of the features and use the traditional clustering algorithm to obtain one clustering result. Repeat the above process t times and obtain various clustering results $\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_t$. The generation process of base clustering results can be depicted as Algorithm 1.

Algorithm 1: Generation of base clustering results.

Input: Data set $U = \{v_1, v_2, \cdots, v_n\}$, ensemble size t, cluster number K
Output: $\mathbb{C}_1^*, \mathbb{C}_2^*, \cdots, \mathbb{C}_t^*$
1 **For** $i = 1 : t$ **do**
2   Randomly extract parts of the features;
3   Use the traditional clustering algorithm to obtain a clustering result $\mathbb{C}_i^*$;
4 **End**
5 **Return** $\mathbb{C}_1^*, \mathbb{C}_2^*, \cdots, \mathbb{C}_t^*$.

3.2. Labels Matching

Based on the base clustering results, we use the voting method to obtain the core region and the fringe region of three way clustering. The base clustering results $\mathbb{C}_1^*, \mathbb{C}_2^*, \cdots, \mathbb{C}_t^*$ cannot be directly used for voting due to the lack of a priori category information. As an example, we consider the dataset $U = (v_1, v_2, v_3, v_4, v_5, v_6)$ and let $\mathbb{C}_1, \mathbb{C}_2$ and $\mathbb{C}_3$ be three clustering results of V, which are shown in Table 1.

Table 1. Different represents of the same clustering results.

	$\mathbb{C}_1$	$\mathbb{C}_2$	$\mathbb{C}_3$
v_1	1	2	3
v_2	1	2	3
v_3	2	3	2
v_4	2	3	2
v_5	3	1	1
v_6	3	1	1

Although the clustering results are expressed in different orders, they represent the same clustering result. In order to combine the clustering results, the cluster labels must be matched to establish the correspondence between each other. Zhou and Tang [31] pointed out that the number of common elements covered by two clusters with the corresponding relationship should be the largest. Therefore, for two base clustering results $\mathbb{C}_i = \{C_i^1, \cdots, C_i^k, \cdots, C_i^K\}$, $\mathbb{C}_j = \{C_j^1, \cdots, C_j^k, \cdots, C_j^K\}$, $1 \leq i, j \leq t$, t is the number of base clusters; we record the number of identical elements covered by each cluster $C_i^{k_1}$, $C_j^{k_2}$ ($1 \leq k_1, k_2 \leq K$) in the two partitions in the OVERLAP matrix of $K \times K$. Then select the cluster label that covers the largest number of identical elements to establish a corresponding relationship, and delete the result from the matrix OVERLAP. Repeat the above process until all cluster labels have established corresponding relationships. This process is defined as label matching. When there are $t(t > 2)$ clustering results, we can select the first clustering results as the matching criterion and match the other clustering results with the first one. The procedure of label matching can be shown as Algorithm 2.

Algorithm 2: Label matching.

   Input: $\mathbb{C}_1^*, \mathbb{C}_2^*, \cdots, \mathbb{C}_t^*$

   Output: $\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_t$

1 **For** $j = 2 : t$ **do**

2   Compute the $K \times K$ OVERLAP matrix between $\mathbb{C}_1^*$ and $\mathbb{C}_j^*$

3   **For** $i = 1{:}K$ **do**

4   Find the largest element of i-row in OVERLAP matrix;

5   Record the corresponding column index q_i of the largest element;

6   Adjust the q_i-th cluster of $\mathbb{C}_j^*$ to i-th cluster;

7 $i = i + 1$;

8   **End**

9 $j = j + 1$;

10 **End**

11 **Return** $\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_t$.

3.3. Results of Three-Way Clustering

After the process of label matching, updated cluster results $\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_t$ can be obtained. Then we use the voting method to achieve the core region and fringe region of three-way clustering. In the voting process, each clustering result can be regarded as a voter, and they will vote once and only once for each data point. Let $C^j = \cup_{i=1}^t C_i^j, (j = 1, 2, \cdots, K)$. For any $v \in C^j$, we count the votes of the element $v \in C_i^j, (i = 1, 2, \cdots, t)$ and denote the votes of the element v as count(v). Suppose p is one given threshold. If count$(v) \geq p$, we assign the element v to the core region of the cluster C^j, otherwise, we assign the element v to the fringe region of the cluster C^j. Finally, we can obtain three-way clustering results. The process of finding the core region and the fringe region can be depicted as Algorithm 3.

Algorithm 3: Finding core region and fringe region of TWCE.

   Input: $\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_t$, threshold p

   Output: $\mathbb{C} = \{(\text{Co}(C^1), \text{Fr}(C^1)), (\text{Co}(C^2), \text{Fr}(C^2)), \cdots, (\text{Co}(C^K), \text{Fr}(C^K))\}$

1 **For** $j = 1 : k$ **do**

2   Compute $\cup_{i=1}^t C_i^j$;

3   **For** $v \in \cup_{i=1}^t C_i^j$ **do**

4   get the vote of the element v: count(v);

5   **If** *count(v)* $\geq p$

6   Assign the element v to the core region of the cluster C_i;

7   **Else**

8   Assign the element v to the fringe region of the cluster C_i.

9   **End**

10   **End**

11 **End**

12 **Return** $\{(\text{Co}(C^1), \text{Fr}(C^1)), (\text{Co}(C^2), \text{Fr}(C^2)), \cdots, (\text{Co}(C^K), \text{Fr}(C^K))\}$.

Based on Algorithms 1–3, a three-way clustering result that describes the data distribution will be generated. Sequentially executing Algorithms 1–3 forms the framework of the proposed TWCE, which is shown as Algorithm 4.

Algorithm 4: Three way clustering based on ensemble strategy (TWCE).

 Input: Data set $U = \{v_1, v_2, \cdots, v_n\}$, ensemble size t, cluster number K,
  threshold p

 Output: $\mathbb{C} = \{(\mathrm{Co}(C^1), \mathrm{Fr}(C^1)), (\mathrm{Co}(C^2), \mathrm{Fr}(C^2)), \cdots, (\mathrm{Co}(C^K), \mathrm{Fr}(C^K))\}$

 1 $\mathbb{C}_1^*, \mathbb{C}_2^*, \cdots, \mathbb{C}_t^* \leftarrow$ Algorithm 1;

 2 $\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_t \leftarrow$ Algorithm 2;

 3 $\{(\mathrm{Co}(C^1), \mathrm{Fr}(C^1)), (\mathrm{Co}(C^2), \mathrm{Fr}(C^2)), \cdots, (\mathrm{Co}(C^K), \mathrm{Fr}(C^K))\} \leftarrow$ Algorithm 3;

 4 **Return** $\mathbb{C} = \{(\mathrm{Co}(C^1), \mathrm{Fr}(C^1)), (\mathrm{Co}(C^2), \mathrm{Fr}(C^2)), \cdots, (\mathrm{Co}(C^K), \mathrm{Fr}(C^K))\}$.

---

4. Experimental Analyses

In this section, we verify the performances of the proposed TWCE strategy. In the process of the experiments, K-means [16] and NJW [63] are used to generate the base clustering results, respectively, and the percentage of the selected feature subsets are randomly 50%, 60%, 70%, 80% or 90%. This section consists of three parts. In the first part, we introduce some popular clustering evaluation indices. In the second part, 12 datasets from the UCI Machine Learning repository are employed to show the working mechanism of the TWCE strategy. The relation between the clustering performances and the percentage of the selected feature subsets is discussed in the third part.

4.1. Evaluation Indices

The evaluation of clustering is an effective process for assessing the performance of clustering algorithms. We compare the proposed algorithm with other existing clustering algorithms by calculating some evaluation indices such as Normalized Mutual Information (NMI) [64], Adjusted Rand Index (ARI) [65] and Accuracy (Acc) [50]. The three validity metrics NMI, ARI and Acc are all positive indices, that is, the larger the value, the better the clustering effect.

1. Normalized Mutual Information (NMI)

$$NMI = \frac{I(X,Y)}{\sqrt{H(X)H(Y)}}, \tag{6}$$

where X is the test label, and Y is the real label. $H(X)$ and $H(Y)$ represent the entropy of X and Y, respectively. $I(X,Y)$ is the mutual information between X and Y.

2. Adjusted Rand Index(ARI)

$$ARI = \frac{2(ad - bc)}{(a+b)(b+d) + (a+c)(c+d)}, \tag{7}$$

where a is the number of data points in a pair that belong to the same cluster in both real and experimental situations; b is the number of data points in a pair that belong to the same cluster in real but not in experimental situations; c is the number of data points in a pair that belong to the same cluster in experimental but not in real situations; d is the number of data points in a pair that do not belong to the same cluster in both real and experimental situations.

3. Accuracy (Acc)

$$Acc = \frac{1}{N} \sum_{i=1}^{k} n_i, \tag{8}$$

where N is the total number of elements, n_i is the number of elements which are correctly divided into corresponding cluster i, and k is cluster number. Acc represents the ratio between the number of correctly partitioned elements and the total number. A higher value of Acc means the clustering result is better.

4.2. Performances of TWCE Strategy

To test the performances of our proposed TWCE strategy, we employ 12 datasets from the UCI Machine Learning repository, which are Cardiotocography, Congressional voting, Dermatology, Forrest, Landsat, Optical recognition, Synthetic, Urban Land Cover, Vehicle, Waveform, Wdbc and Wine. Table 2 shows the details of these datasets. The first step of the TWCE strategy is to obtain base clustering results. Different clustering algorithms may obtain different clustering results. We use the K-means algorithm and the NJW algorithm to generate base clusters in our experiments, which randomly select 50%, 60%, 70%, 80% or 90% feature subsets to clustering, respectively.

Table 2. A description of dataset used.

ID	Data Sets	Samples	Attributes	Classes
1	Cardiotocography	2126	21	10
2	Congressional voting	435	16	2
3	Dermatology	366	34	6
4	Forrest	523	27	4
5	Landsat	6435	36	6
6	Optical recognition	5620	64	10
7	Synthetic	600	60	6
8	Urban Land Cover	675	147	9
9	Vehicle	846	18	4
10	Waveform	5000	21	3
11	Wdbc	569	30	2
12	Wine	178	13	3

Because the evaluation indices NMI, ARI and Acc are only adopted to the hard clustering results, a three-way clustering results cannot calculate these values directly. In order to present the performances of our proposed TWCE algorithm, we use the core regions to form a clustering result, then calculate the NMI, ARI and Acc valuse by using the core region to represent the corresponding cluster. The average NMI value, ARI value and Acc value are achieved by running 30 times on all datasets and the ensemble size is 50. Tables 3 and 4 show the performances of the TWCE strategy based on the K-means and NJW algorithm, respectively. For comparing the clustering effect, the performances of K-means, NJW, Voting [31] and CSPA [29] are also presented in Tables 5–7.

Table 3. The performances of TWCE using K-means.

ID	Data Sets	NMI	ARI	Acc
1	Cardiotocography	0.6795	0.5746	0.7611
2	Congressional voting	0.5648	0.6475	0.9026
3	Dermatology	0.9777	0.9776	0.9748
4	Forrest	0.6040	0.5709	0.8221
5	Landsat	0.7151	0.6491	0.7829
6	Optical recognition	0.9573	0.9654	0.9789
7	Synthetic	0.9008	0.8761	0.8947
8	Urban Land Cover	0.6851	0.6543	0.8233
9	Vehicle	0.2812	0.1888	0.4792
10	Waveform	0.4478	0.3108	0.5574
11	Wdbc	0.6875	0.7763	0.9412
12	Wine	0.9219	0.9418	0.9800

Table 4. The performances of TWCE using NJW.

ID	Data Sets	NMI	ARI	Acc
1	Cardiotocography	0.6320	0.5135	0.7721
2	Congressional voting	0.5619	0.6430	0.9012
3	Dermatology	0.9850	0.9902	0.9914
4	Forrest	0.6446	0.6410	0.8525
5	Landsat	0.7776	0.7444	0.8440
6	Optical recognition	0.9623	0.9700	0.9798
7	Synthetic	0.9121	0.8885	0.9053
8	Urban Land Cover	0.6870	0.6380	0.8059
9	Vehicle	0.2648	0.1874	0.5081
10	Waveform	0.3326	0.1462	0.5361
11	Wdbc	0.6835	0.7920	0.9452
12	Wine	0.9230	0.9437	0.9812

Table 5. The performances of average NMI value.

ID	Data Sets	K-Means	NJW	Voting	CSPA
1	Cardiotocography	0.3305	0.3314	0.3772	0.3256
2	Congressional voting	0.4574	0.4462	0.4541	0.4640
3	Dermatology	0.8130	0.8226	0.9133	0.8614
4	Forrest	0.5424	0.5428	0.5439	0.4898
5	Landsat	0.6125	0.6028	0.6482	0.6134
6	Optical recognition	0.7335	0.7312	0.7750	0.7385
7	Synthetic	0.7418	0.7587	0.8035	0.7292
8	Urban Land Cover	0.5770	0.5239	0.6029	0.5491
9	Vehicle	0.1126	0.0844	0.1418	0.1058
10	Waveform	0.3642	0.3668	0.2682	0.3639
11	Wdbc	0.6232	0.6075	0.6190	0.6175
12	Wine	0.8249	0.8782	0.8369	0.8354

Table 6. The performances of average ARI value.

ID	Data Sets	K-Means	NJW	Voting	CSPA
1	Cardiotocography	0.1611	0.1547	0.2700	0.1567
2	Congressional voting	0.5287	0.5263	0.5381	0.5368
3	Dermatology	0.6544	0.7487	0.8446	0.7685
4	Forrest	0.4956	0.4904	0.4952	0.4299
5	Landsat	0.5264	0.4961	0.5725	0.5276
6	Optical recognition	0.6358	0.6583	0.7215	0.6525
7	Synthetic	0.5902	0.6336	0.7135	0.5730
8	Urban Land Cover	0.4587	0.4021	0.5245	0.4181
9	Vehicle	0.0800	0.0657	0.0965	0.0829
10	Waveform	0.2535	0.2496	0.2178	0.2524
11	Wdbc	0.7302	0.7251	0.7259	0.7242
12	Wine	0.8367	0.8992	0.8564	0.8449

Table 7. The performances of average *Acc* value.

ID	Data Sets	K-Means	NJW	Voting	CSPA
1	Cardiotocography	0.3701	0.3384	0.5484	0.3653
2	Congressional voting	0.8610	0.8631	0.8671	0.8667
3	Dermatology	0.6996	0.8077	0.8508	0.8117
4	Forrest	0.7795	0.7540	0.7807	0.7130
5	Landsat	0.6682	0.6827	0.7429	0.6707
6	Optical recognition	0.7556	0.7716	0.8450	0.7654
7	Synthetic	0.6685	0.7475	0.8337	0.6697
8	Urban Land Cover	0.6119	0.5838	0.7333	0.5910
9	Vehicle	0.3691	0.3738	0.4243	0.3845
10	Waveform	0.5011	0.5049	0.5607	0.5013
11	Wdbc	0.9279	0.9262	0.9267	0.9262
12	Wine	0.9409	0.9663	0.9517	0.9292

From the experimental results in Tables 3–7, we can make the following conclusions.

(1) It is obvious that the *NMI* and *ARI* performances of the TWCE strategy based on K-means and NJW are better than only using K-means and NJW. So, the TWCE strategy can indeed obtain a better clustering result than a single clustering algorithm. Compared to the other two ensemble clustering algorithms, the *NMI* value and *ARI* value obtained by the TWCE algorithm outperform on most of the 12 datasets. Specifically, the performances of TWCE based on K-means are always superior to the other algorithms on all datasets, and the performances of TWCE based on NJW are only slightly worse than the voting algorithms on Waveform. The improvement can be attributed to the fact that the clustering result of the proposed TWCE algorithm is represented by core regions when calculating the *NMI* value and the *ARI* value, which can increase the degree of separation between clusters and reduce the degree of dispersion within clusters.

(2) TWCE based on K-means or NJW both have a better *Acc* value on most datasets compared to other algorithms, except Waveform. The increase of *Acc* value can be attributed to each cluster being represented only by a core region when we calculate the *Acc* value, which means that the total number is to exclude elements in the fringe region, that is, n_i and N both become smaller in Equation (3).

(3) Comparing the performances of TWCE using K-means and TWCE using NJW, we can find that using different clustering algorithms to generate base clusters has little effect on the performances of the TWCE algorithm.

4.3. The Influences of the Selected Feature Percentage

The TWCE strategy uses a traditional clustering algorithm on parts of the features to obtain the base clustering results. The percentage of the selected feature subsets will have an impact on the performances of the experiment. In this subsection, we discuss the relation between the clustering performances and the percentage of the selected feature subsets. We also use the datasets in the previous subsection and the percentages of the feature subset are set to 50%, 60%, 70%, 80% and 90% of a dataset when the ensemble size is 50. The average *NMI* value, *ARI* value and *Acc* value are achieved by running the TWCE algorithm 30 times, which uses the K-means algorithm and the NJW algorithm to generate base clusters on all datasets, respectively. Figures 2–13 list the *NMI* value, *ARI* value and *Acc* value of twelve datasets from the TWCE strategy based on K-means when the percentage of the feature subset takes different values. Figures 14–25 show the relation between the clustering performances and the percentage of the feature subsets selected by the TWCE strategy based on the NJW algorithm.
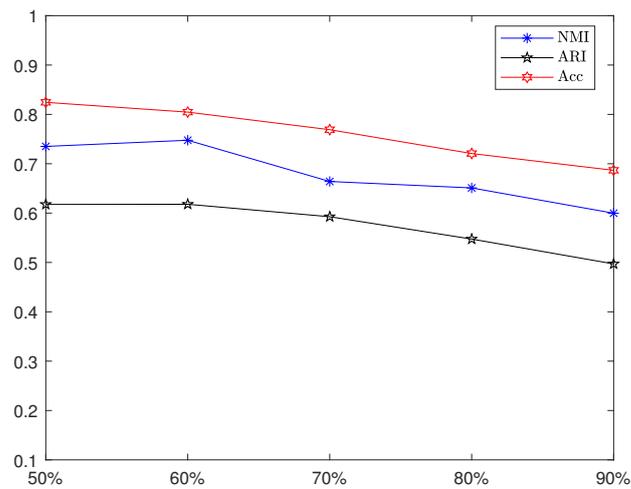
Figure 2. Results of Cardiotocography by TWCE strategy based on K-means.
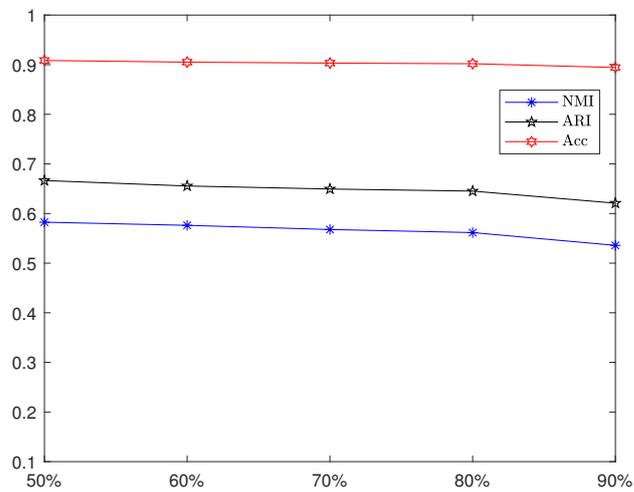

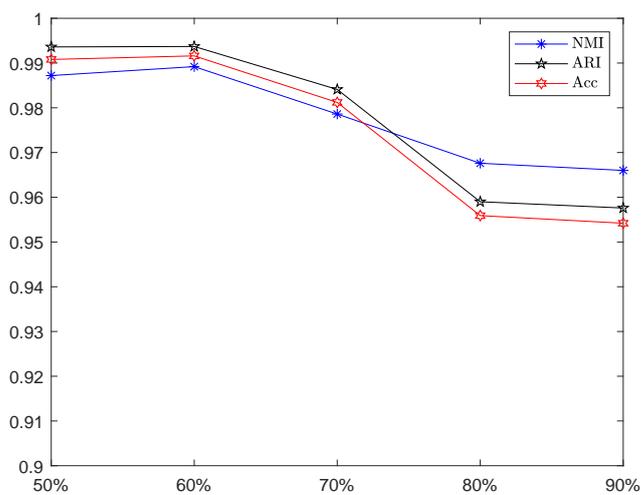**Figure 3.** Results of Congressional Voting by TWCE strategy based on K-means.

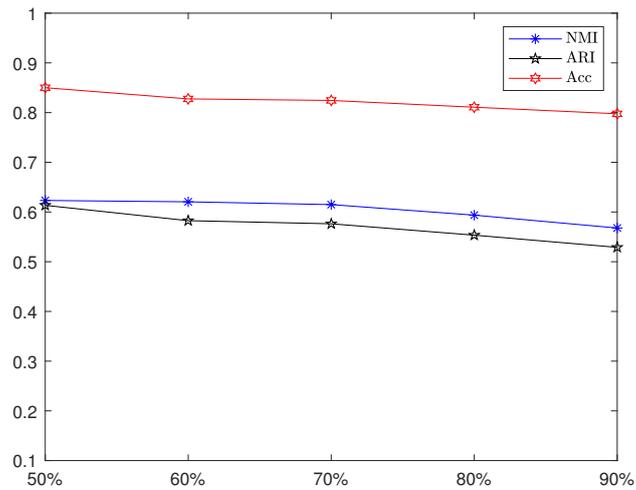**Figure 4.** Results of Dermatology by TWCE strategy based on K-means.

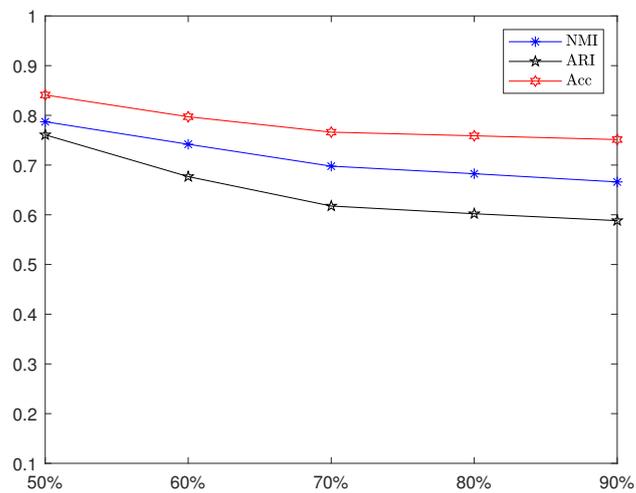**Figure 5.** Results of Forrest by TWCE strategy based on K-means.

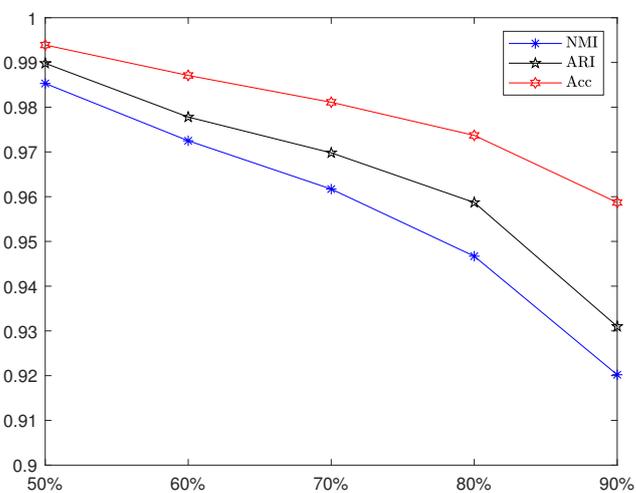**Figure 6.** Results of Landsat by TWCE strategy based on K-means.

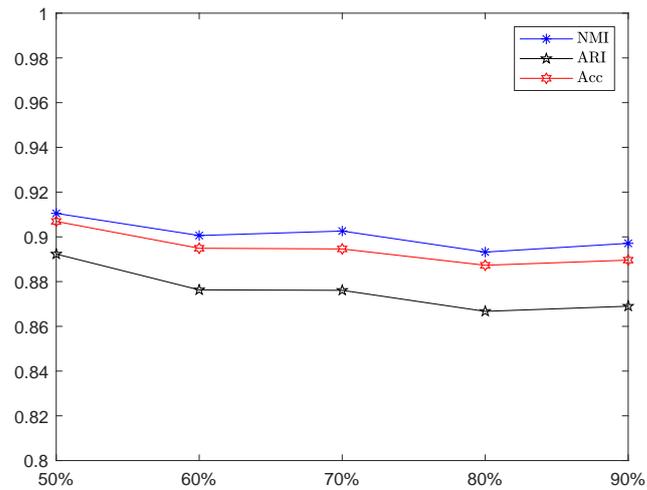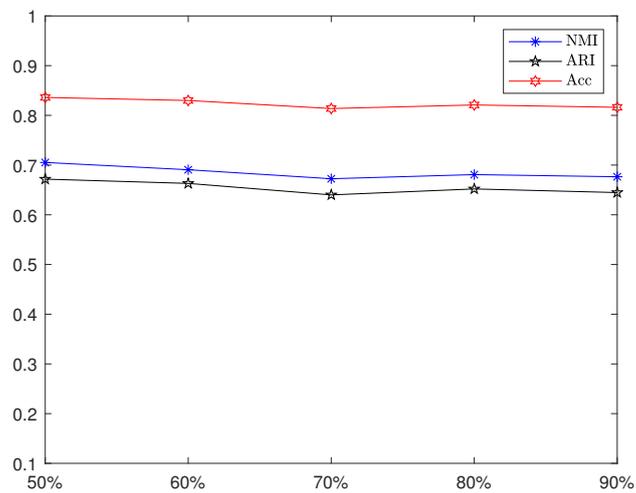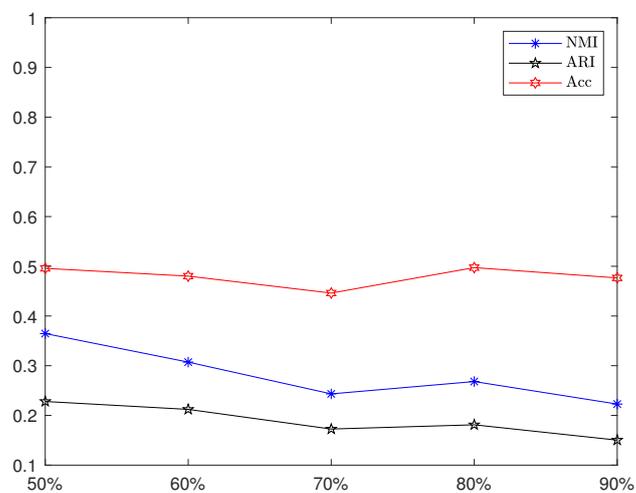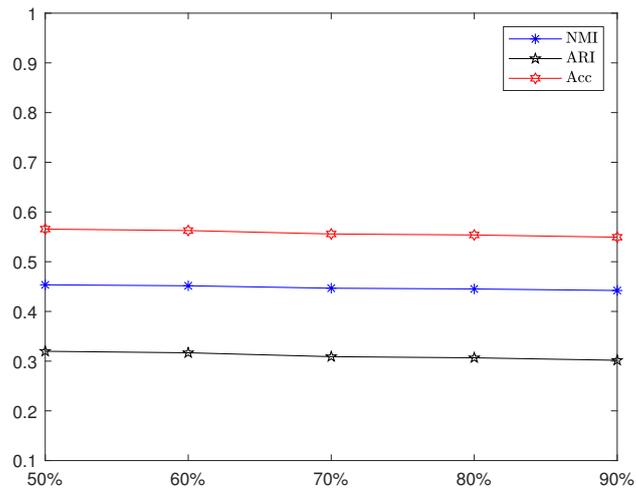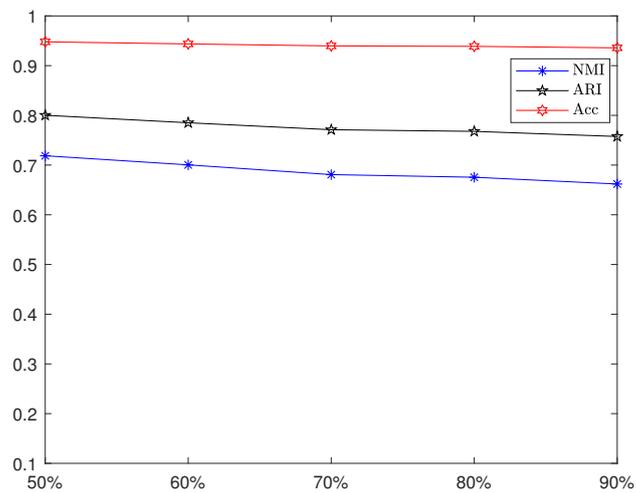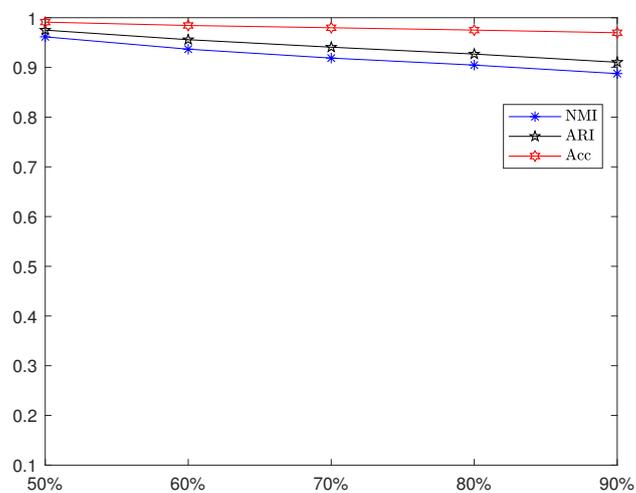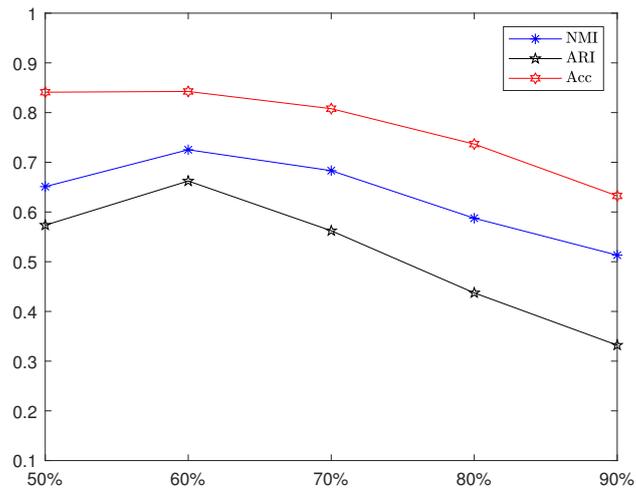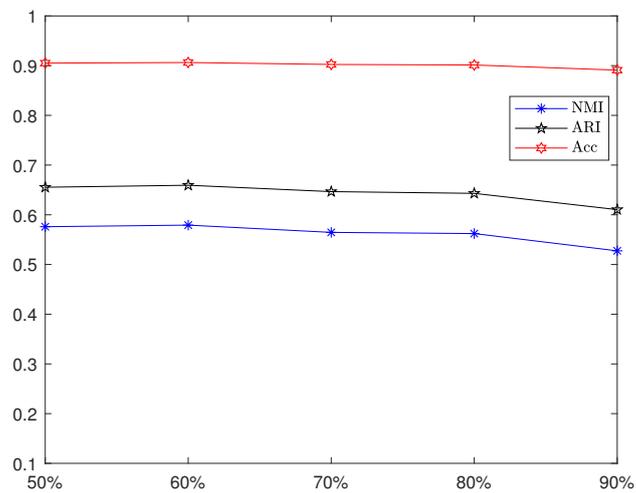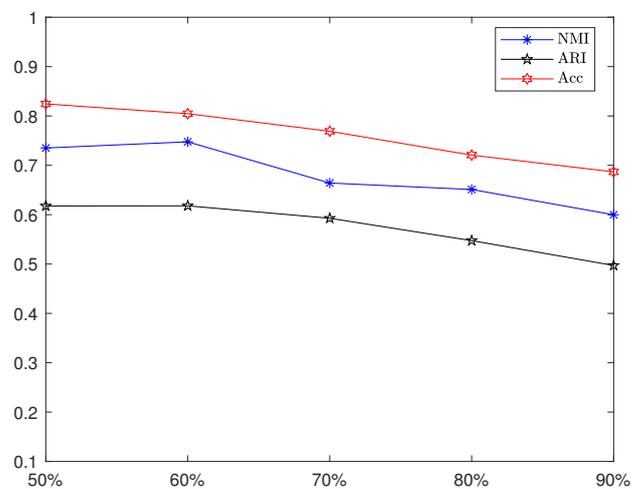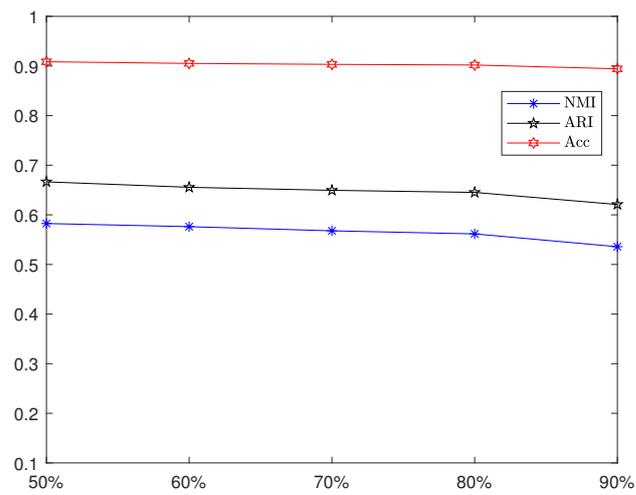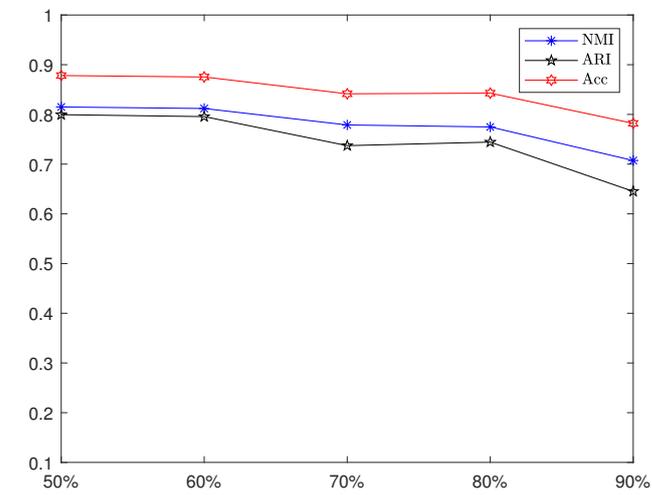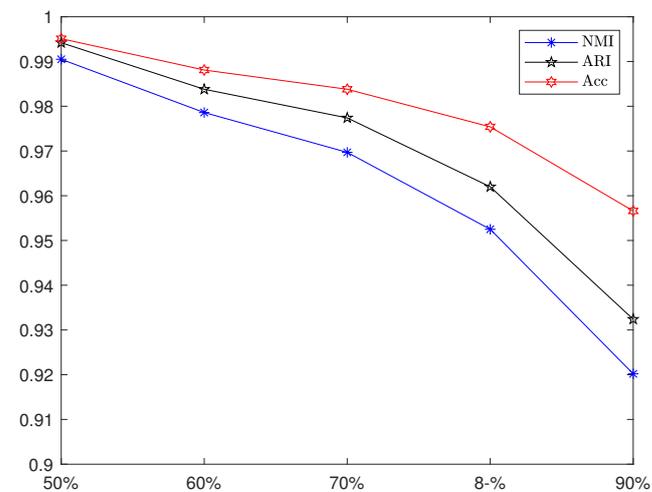**Figure 7.** Results of Optical recognition by TWCE strategy based on K-means.
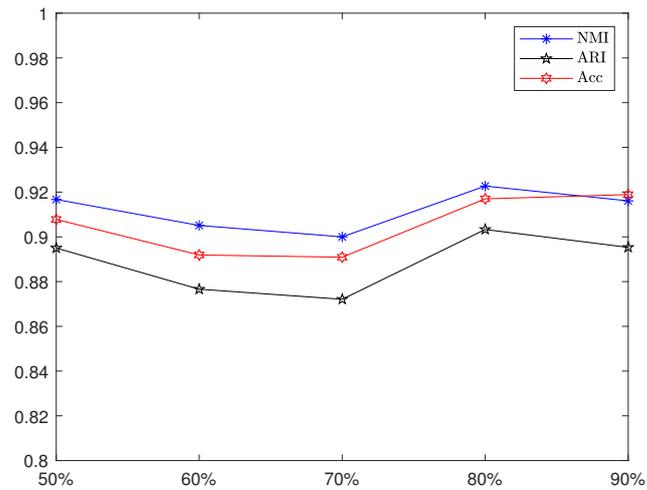
Figure 8. Results of Synthetic by TWCE strategy based on K-means.

Figure 9. Results of Urban Land Cover by TWCE strategy based on K-means.

Figure 10. Results of Vehicle by TWCE strategy based on K-means.

Figure 11. Results of Waveform by TWCE strategy based on K-means.

Figure 12. Results of Wdbc by TWCE strategy based on K-means.

Figure 13. Results of Wine by TWCE strategy based on K-means.

Figure 14. Results of Cardiotocography by TWCE strategy based on NJW.

Figure 15. Results of Congressional Voting by TWCE strategy based on NJW.

Figure 16. Results of Dermatology by TWCE strategy based on NJW.

Figure 17. Results of Forrest by TWCE strategy based on NJW.

Figure 18. Results of Landsat by TWCE strategy based on NJW.

Figure 19. Results of Optical recognition by TWCE strategy based on NJW.

Figure 20. Results of Synthetic by TWCE strategy based on NJW.

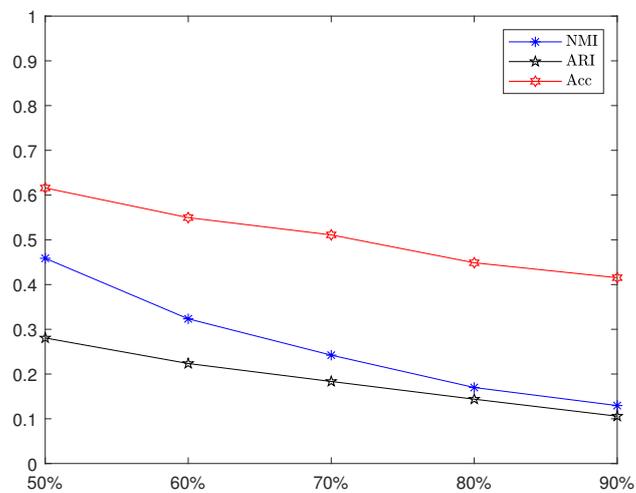**Figure 21.** Results of Urban Land Cover by TWCE strategy based on NJW.

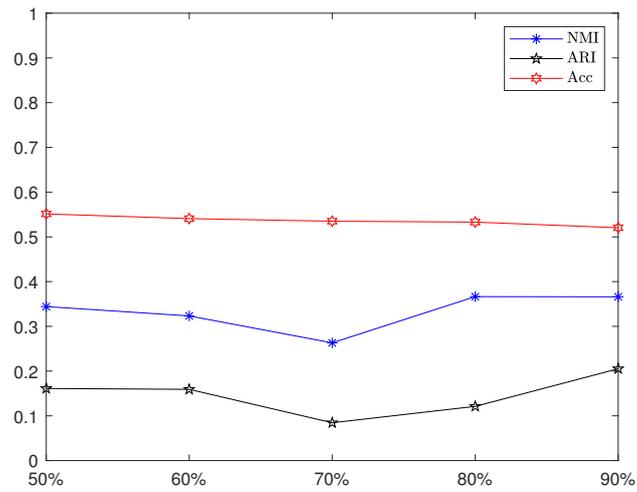**Figure 22.** Results of Vehicle by TWCE strategy based on NJW.

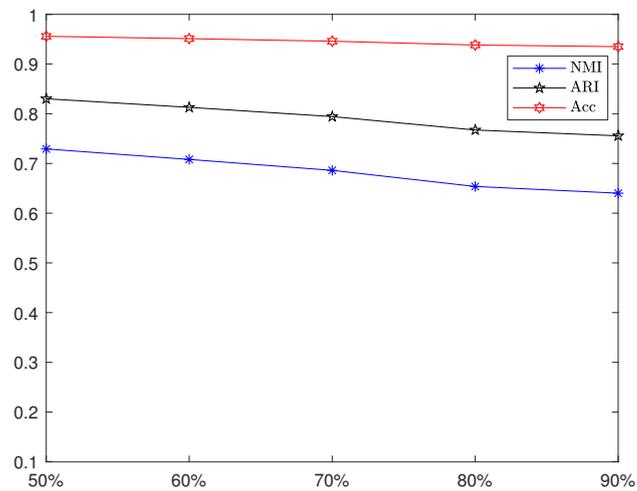**Figure 23.** Results of Waveform by TWCE strategy based on NJW.

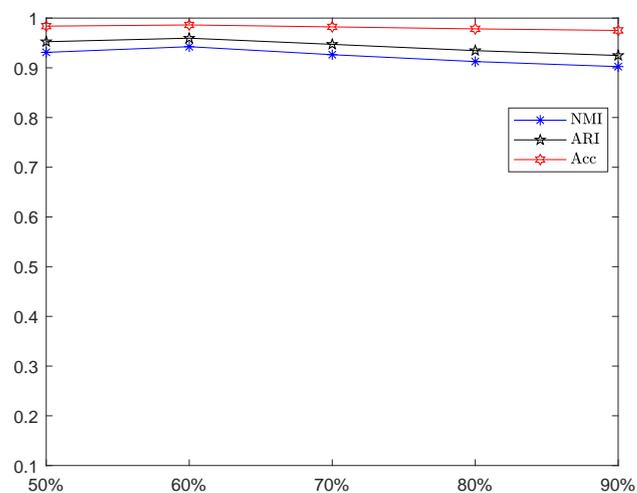**Figure 24.** Results of Wdbc by TWCE strategy based on NJW.

Figure 25. Results of Wine by TWCE strategy based on NJW.

From the experimental results recorded in Figures 2–25, we can find that different datasets achieve the best performances at different percentages. For example, Dermatology achieves the best performances with the TWCE strategy based on K-means when the percentage is 60%, while most of the other datasets achieve the best performances with the TWCE strategy based on K-means when the percentage is 50%. Even for the same dataset, the TWCE strategy on different clustering algorithms achieves the best performances at different percentages. For example, Synthetic achieves the best performance with the TWCE strategy based on K-means when the percentage is 70%, while the best performance by the TWCE strategy based on NJW is obtained at a different percentage value. Though different datasets achieve the best performances at different percentages, the experimental performances of most of the datasets become worse when the percentage is 90%. This is because the diversity of the base clustering results will become smaller when the percentage becomes larger and low diversity limits the improvement of the ensemble performance. The issue of choosing a reasonable percentage needs to be further explored in future research.

5. Conclusions and Future Work

It has been recognized that a single clustering algorithm cannot identify all types of data structures. Ensemble clustering is an effective approach to solving the problem that a single clustering algorithm may not obtain good clustering results for all datasets. The three-way clustering method uses a core region and a fringe region to solve the problem of inaccurate decision-making caused by inaccurate information or insufficient data. Integrating the idea of three-way clustering and ensemble clustering methods, we propose a new ensemble three-way clustering strategy in this paper. In the proposed strategy, we randomly extract part of the features and use the traditional clustering algorithm to obtain one clustering result. A different feature subset will lead to different clustering results. Diverse base clustering results can be obtained by using different feature subsets. Based on the base clustering results, we use label matching to align all clustering results in a given order and the voting method to obtain the core region and the fringe region of the three way clustering. The sample is assigned to the core region of the corresponding cluster when the frequency of the sample in the same cluster is more than the given threshold. The difference between the union of the cluster with the same labels and the core region is regarded as the fringe region of the specific cluster. Therefore, a three-way clustering strategy is obtained. As examples for demonstration, we apply the proposed strategy on the top of K-means and spectral clustering, respectively. The experimental results on UCI datasets demonstrate the effectiveness in revealing data structures.

The following topics will deserve further investigation:

(1) In this paper, the cluster number *K* is set to be constant during the process of generating base cluster members. Due to clustering, the algorithm is an unsupervised method, so how to apply the proposed algorithm to the different *K* is our next future work.

(2) The base clusters generated by different feature subsets may be of low quality which may effect the final ensemble clustering result. We can evaluate the quality of base clusters by setting the evaluation function to remove some low-quality members of base clusters. This will be a good research direction.

(3) In the process of three-way decision, the strategy needs to obtain more details on the division guidelines, so that the proposed algorithm can achieve a clustering result with a higher performance.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: http://archive.ics.uci.edu/ml/datasets (accessed on 21 March 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ding, S.F.; Jia, H.J.; Du, M.J.; Xue, Y. A semi-supervised approximate spectral clustering algorithm based on HMRF model. *Inf. Sci.* **2018**, *429*, 215–228. [CrossRef]
2. Shi, H.; Wang, P.X.; Yang, X.B.; Yu, H.L. An improved mean imputation clustering algorithm for incomplete data. *Neural Process. Lett.* **2021**. [CrossRef]
3. Yang, X.B.; Qi, Y.S.; Song, X.N.; Yang, J.Y. Test cost sensitive multigranulation rough set: Model and minimal cost selection. *Inf. Sci.* **2013**, *250*, 184–199. [CrossRef]
4. Xu, W.H.; Guo, Y.T. Generalized multigranulation double-quantitative decision-theoretic rough set. *Knowl.-Based Syst.* **2016**, *105*, 190–205. [CrossRef]
5. Li, W.T.; Xu, W.H.; Zhang, X.Y.; Zhang, J. Updating approximations with dynamic objects based on local multigranulation rough sets in ordered information systems. *Artif. Intell. Rev.* **2021**, *55*, 1821–1855. [CrossRef]
6. Xu, W.H.; Yu, J.H. A novel approach to information fusion in multi-source datasets: A granular computing viewpoint. *Inf. Sci.* **2017**, *378*, 410-423. [CrossRef]
7. Chen, X.W.; Xu, W.H. Double-quantitative multigranulation rough fuzzy set based on logical operations in multi-source decision systems. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 1021–1048. [CrossRef]
8. Xu, W.H.; Yuan, K.H.; Li, W.T. Dynamic updating approximations of local generalized multigranulation neighborhood rough set. *Appl. Intell.* **2022**. [CrossRef]
9. Yang, X.B.; Yao, Y.Y. Ensemble selector for attribute reduction. *Appl. Soft Comput.* **2018**, *70*, 1–11. [CrossRef]
10. Jiang, Z.; Yang, X.; Yu, H., Liu, D.; Wang, P.; Qian, Y. Accelerator for multi-granularity attribute reduction. *Knowl.-Based Syst.* **2019**, *177*, 145–158. [CrossRef]
11. Li, J.; Yang, X.; Song, X.; Li, J.; Wang, P.; Yu, D.J. Neighborhood attribute reduction: A multi-criterion approach. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 731–742. [CrossRef]
12. Liu, K.; Yang, X.; Yu, H.; Fujita, H.; Chen, X.; Liu, D. Supervised information granulation strategy for attribute reduction. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 2149–2163. [CrossRef]
13. Xu, S.; Yang, X.; Yu, H.; Yu, D.J.; Yang, J.; Tsang, E.C. Multi-label learning with label-specific feature reduction. *Knowl.-Based Syst.* **2016**, *104*, 52–61. [CrossRef]
14. Liu, K.Y.; Yang, X.B.; Fujita, H. An efficient selector for multi-granularity attribute reduction. *Inf. Sci.* **2019**, *505*, 457–472. [CrossRef]
15. Liu, K.; Yang, X.; Yu, H.; Mi, J.; Wang, P.; Chen, X. Rough set based semi-supervised feature selection via ensemble selector. *Knowl.-Based Syst.* **2020**, *165*, 282–296. [CrossRef]
16. MacQueen J. Some methods for classification and analysis of multivariate observations. *Berkeley Symp. Math. Stat. Probab.* **1967**, *5*, 281–297.
17. Maulik, U.; Bandyopadhyay, S. Genetic algorithm-based clustering technique. *Pattern Recognit.* **2000**, *33*, 1455–1465. [CrossRef]
18. Gurrutxaga, I.; Albisua, I.; Arbelaitz, O.; Martín, J.I.; Muguerza, J.; Pérez, J.M.; Perona, I. An efficient method to find the best partition in hierarchical clustering based on a new cluster validity index. *Pattern Recognit.* **2010**, *43*, 3364–3373. [CrossRef]
19. Fred A L, Leito M N. Partitional vs. hierarchical clustering using a minimum grammar complexity approach. In Proceeding of the SSPR 2000&SPR 2000, Alicante, Spain, 30 August–1 September 2000; pp. 193–202.
20. Azzag, H.; Lebbah, M. *A New Way for Hierarchical and Topological Clustering*; Guillet, F., Pinaud, B., Venturini, G., Zighed, D.; Eds.; Advances in Knowledge Discovery and Management; Springer: Berlin/Heidelberg, Germany, 2013; pp. 85–97.
21. Lotfi, A; Moradi, P.; Beigy, H. Density peaks clustering based on density backbone and fuzzy neighborhood. *Pattern Recognit.* **2020**, *107*, 107449. [CrossRef]
22. Ankerts, M.; Breuning, M.M.; Kriegel, H.P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. International Conference on Management of Data and Symposium on Principles of Database Systems, Philadelphia, PA, USA, 31 May–3 June 1999; pp. 49–60.
23. Hinneburg, A.; Keim, D.A. An efficient approach to clustering in large multimedia databases with noise. In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 58–65.
24. Birant, D.; Kut, A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [CrossRef]
25. Agrawal, R.; Gehrke, J.; Gunopulos, D.; Raghavan, P. *Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications*; ACM: New York, NY, USA, 1998.
26. Govaert, G.; Nadif, M. An EM algorithm for the block mixture model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 643–647. [CrossRef] [PubMed]

27. Pavlidis, P.; Li, Q.; Noble, W.S. The effect of replication on gene expression microarray experiments. *Bioinformatics* **2003**, *19*, 1620–1627. [CrossRef] [PubMed]
28. Yen, L.C.; Lu, M.C.; Chen, J.L. Applying the self-organization feature map (SOM) algorithm to AE-based tool wear monitoring in micro-cutting. *Mech. Syst. Signal Process.* **2013**, *34*, 353–366. [CrossRef]
29. Strehl, A.; Ghosh, J. Cluster ensembles-a knowledge reuse framework for combing multiple partitions. *J. Mach. Learn. Res.* **2003**, *3*, 583–617.
30. Fred, A.L.N.; Jain, AK. Combining multiple clusterings using evidence accumulation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 835–850. [CrossRef]
31. Zhou, Z.H.; Tang, W. Cluster Ensemble. *Knowl.-Based Syst.* **2006**, *19*, 77–83. [CrossRef]
32. Huang, D.; Lai, J.H.; Wang, C.D. Ensemble clustering using factor graph. *Pattern Recognit.* **2016**, *50*, 131–142. [CrossRef]
33. Huang, D.; Wang, C.D.; Lai, J.H. Locally weighted ensemble clustering. *IEEE Trans. Cybern.* **2018**, *48*, 1460–1473. [CrossRef]
34. Xu, L.; Ding, S.F. A novel clustering ensemble model based on granular computing. *Appl. Intell.* **2021**, *51*, 5474–5488. [CrossRef]
35. Zhou, P.; Wang, X.; Du, L.; Li, X.J. Clustering ensemble via structured hypergraph learning. *Inf. Fusion* **2022**, *78*, 171–178. [CrossRef]
36. Yao, Y.Y. The superiority of three-way decisions in probabilistic rough set models. *Inf. Sci.* **2011**, *81*, 1080–1096. [CrossRef]
37. Yao, Y.Y. Three-way decisions and cognitive computing. *Cogn. Comput.* **2016**, *8*, 543–554. [CrossRef]
38. Yao, Y.Y. Tri-level thinking: Models of three-way decision. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 947–959. [CrossRef]
39. Yao, Y.Y. The geometry of three-way decision. *Appl. Intell.* **2021**, *51*, 6298–6325. [CrossRef]
40. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 314–356. [CrossRef]
41. Zadeh, LA. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
42. Pedrycz, W. Shadowed sets: Representing and processing fuzzy sets. *IEEE Trans. Syst. Man Cybern. B* **1988**, *28*, 103–109. [CrossRef]
43. Xu, W.H.; Li, W.T. Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets. *IEEE Trans. Cybern.* **2016**, *46*, 366–379. [CrossRef]
44. Yuan, K.H.; Xu, W.H.; Li, W.T.; Ding, W.T. An incremental learning mechanism for object classificationbased on progressive fuzzy three-way concept. *Inf. Sci.* **2022**, *584*, 127–147 [CrossRef]
45. Yu, H. A framework of three-way cluster analysis. In Proceedings of the International Joint Conference on Rough Sets Olsztyn, Poland, 3–7 July 2017; pp. 300–312.
46. Yu, H.; Jiao, P.; Yao, Y.Y.; Wang, G.Y. Detecting and refining overlapping regions in complex networks with three-way decisions. *Inf. Sci.* **2016**, *373*, 21–41. [CrossRef]
47. Yu, H.; Wang, X.C.; Wang, G.Y.; Zeng, X.H. An active three-way clustering method via low-rank matrices for multi-view data. *Inf. Sci.* **2020**, *507*, 823–839. [CrossRef]
48. Afridi, M.K.; Azam, N.; Yao, J.T. A three-way clustering approach for handling missing data using gtrs. *Int. J. Approx. Reason.* **2018**, *98*, 11–24. [CrossRef]
49. Wang, P.X.; Yao, Y.Y. CE3: A three-way clustering method based on mathematical morphology. *Knowl.-Based Syst.* **2018**, *155*, 54–65. [CrossRef]
50. Wang, P.X.; Shi, H.; Yang, X.B.; Mi, J.S. Three-way k-means: Integrating k-means and three-way decision. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2767–2777. [CrossRef]
51. Wang, P.X.; Chen, X.J. Three-way ensemble clustering forincomplete data. *IEEE Access* **2020**, *8*, 91855–91864. [CrossRef]
52. Wang, P.X.; Yang, X.B. Three-way clustering method based on stability theory. *IEEE Access* **2021**, *9*, 33944–33953. [CrossRef]
53. Zhu, J.; Jiang, D.Q.; Wang, P.X. A Three-step Method for Three-way Clustering by Similarity-based Sample's Stability. *Math. Probl. Eng.* **2022**, *2022*, 6555501. [CrossRef]
54. Chu, X.; Sun, B.; Li, X.; Han, K.; Wu, J.; Zhang, Y.; Huang, Q. Neighborhood rough set-based three-way clustering considering attribute correlations: An approach to classification of potential gout groups. *Inf. Sci.* **2020**, *535*, 28–41. [CrossRef]
55. Yu, H.; Chen, Y.; Lingras, P.; Wang, G.Y. A three-way cluster ensemble approach for large-scale data. *Int. J. Approx. Reason.* **2019**, *115*, 32–49. [CrossRef]
56. Shah, A.; Azam, N.; Alanazi, E.; Yao, J.T. Image blurring and sharpening inspired three-way clustering approach. *Appl. Intell.* **2022**. [CrossRef]
57. Wang, X.; Yang, C.Y.; Zhou, J. Clustering aggregation by probability accumulation. *Pattern Recognit.* **2009**, *42*, 668–675. [CrossRef]
58. Punera, K.; Ghosh, J. Consensus-based ensembles of Soft clusterings. *Appl. Artifical Intell.* **2008**, *22*, 780–810. [CrossRef]
59. Sevillano, X.; Alías, F.; Socoró, J.C. Positional and confidence voting-based consensus functions for fuzzy cluster ensembles. *Fuzzy Sets Syst.* **2012**, *193*, 1–32. [CrossRef]
60. Li, F.J.; Qian, Y.H.; Wang, J.T.; Dang, C.Y.; Li, L.P. Clustering ensemble based on sample's stability. *Artif. Intell.* **2019**, *273*, 37–55. [CrossRef]
61. Zhang, K. A three-way c-means algorithm. *Appl. Soft Comput.* **2019**, *82*, 105336. [CrossRef]
62. Jia, X.Y.; Rao, Y.; Li, W.W.; Yang, S.C.; Yu, H. An automatic three-way clustering method based on sample similarity. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 1545–1556. [CrossRef]
63. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On spectral clustering: Analysis and an algorithm. In Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, Vancouver, BC, Canada, 3–8 December 2001; pp. 849–856.

64. Vinh, L.T.; Lee, S.; Park, Y.T.; D'Auriol, B.J. A novel feature selection method based on normalized mutual information. *Appl. Intell.* **2012**, *37*, 100–120. [CrossRef]
65. Lawrence, H.; Phipps, A. Comparing partitions. *J. Classif.* **1985**, *2*, 193–218.