



# An automatic three-way clustering method based on sample similarity

Xiuyi Jia<sup>1</sup> · Ya Rao<sup>1</sup> · Weiwei Li<sup>2</sup> · Sichun Yang<sup>3</sup> · Hong Yu<sup>4</sup>

Received: 17 July 2020 / Accepted: 9 December 2020 / Published online: 14 January 2021  
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

The three-way clustering is an extension of traditional clustering by adding the concept of fringe region, which can effectively solve the problem of inaccurate decision-making caused by inaccurate information or insufficient data in traditional two-way clustering methods. The existing three-way clustering works often select the appropriate number of clusters and the thresholds for three-way partition according to subjective tuning. However, the method of fixing the number of clusters and the thresholds of the partition cannot automatically select the optimal number of clusters and partition thresholds for different data sets with different sizes and densities. To address the above problem, this paper proposed an improved three-way clustering method. First, we define the roughness degree by introducing the sample similarity to measure the uncertainty of the fringe region. Moreover, based on the roughness degree, we define a novel partitioning validity index to measure the clustering partitions and propose an automatic threshold selection method. Second, based on the concept of sample similarity, we introduce the intra-class similarity and the inter-class similarity to describe the quantitative change of the relationship between the sample and the clusters, and define a novel clustering validity index to measure the clustering performance under different numbers of clusters through the integration of the above two kinds of similarities. Furthermore, we propose an automatic cluster number selection method. Finally, we give an automatic three-way clustering approach by combining the proposed threshold selection method and the cluster number selection method. The comparison experiments demonstrate the effectiveness of our proposal.

**Keywords** Three-way decisions · Three-way clustering · Sample similarity

---

This work is supported by the National Natural Science Foundation of China (Grant Nos. 61773208, 61906090, 61876027 and 71671086), the Natural Science Foundation of Jiangsu Province (Grant No. BK20191287), the Natural Science Foundation of Anhui Province of China (Grant No. 1808085MF178), the Fundamental Research Funds for the Central Universities (Grant No. 30920021131), and the China Postdoctoral Science Foundation (Grant No. 2018M632304).

---

✉ Weiwei Li  
liweiwei@nuaa.edu.cn

Xiuyi Jia  
jiaxy@njust.edu.cn

Ya Rao  
Xiaoya\_Rao@163.com

Sichun Yang  
yangsc@ahut.edu.cn

Hong Yu  
yuhong@cqupt.edu.cn

## 1 Introduction

As a research hotspot in the field of rough sets, three-way decisions theory [29] is an extension of classical two-way decisions theory, in which a deferment decision (or non-commitment decision) is adopted when current information is limited or insufficient to make acceptance decision or rejection decision. Three-way decisions is a kind of decision-making theory to describe human beings in dealing

- <sup>1</sup> School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
- <sup>2</sup> College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
- <sup>3</sup> School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China
- <sup>4</sup> College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

with uncertainty information based on the semantic studies of positive rules, boundary rules and negative rules coming from rough set theory. When the decision makers have sufficient confidence based on given information, the acceptance or rejection in two-way decisions are usually adopted to output an immediately result. However, when the decision makers have insufficient confidence in the prejudgment based on current information, the deferment decision is often adopted to reduce the risk brought by the immediate decisions. For uncertainty information, whether it is to accept or reject, there is a high probability of making the wrong decision.

Three-way clustering [32, 35, 36] is an important application of the three-way decisions theory, which can effectively solve the problem of inaccurate partition caused by the incomplete information or insufficient data in the traditional two-way clustering methods. Compared with two-way clustering methods, the three-way clustering incorporates the concept of fringe region (or boundary region) for uncertain samples, and the clustering result is mainly affected by the number of clusters and the thresholds for making three-way decisions. In the existing works, people usually choose the appropriate number of clusters based on the expert opinion, and select the same constant threshold for all data in the iterations of making three-way decisions. However, the choices of such fixed thresholds and number of clusters does not provide a good indication of the differences between clusters or data sets, especially for data sets with different sizes and densities.

Generally, three-way clustering can be regarded as an effective combination of unsupervised clustering and three-way partitioning of clusters [33]. Firstly, from the perspective of clustering validity, it is expected that the samples in different clusters after clustering are easier to distinguish, that is, the distance between different clusters is as far as possible. For the samples in the same cluster, it is expected that the samples are as close as possible, that is, the smaller the distance between the samples within the cluster, the better [6]. Similarly, for the three-way partitioning of clusters, it is to improve the accuracy of the samples contained in each cluster by introducing the fringe region, which is beneficial for the accuracy of the overall clustering result [32]. However, the introduction of the fringe region also increases the uncertainty of knowledge representation of the clustering result. Therefore, from the perspective of partition validity, the sample distribution and its uncertainty in the fringe region are the key factors affecting the performance of the cluster partitioning. In view of this, for three-way clustering, it is expected to improve the performance of clustering by selecting appropriate number of clusters and thresholds for three-way partitioning. To be specific, appropriate number of clusters should be selected to make the samples within same cluster closer and the samples in different clusters farther, and appropriate thresholds for three-way partitioning

should be selected to maximize the accuracy of the samples contained in each cluster and minimize the uncertainty of knowledge representation of the final clustering result.

To address the above problems, in this paper, we propose an automatic three-way clustering method based on sample similarity, in which the quantitative changes of the relationship between all samples and clusters are considered. First, we define sample similarity to describe the quantitative relationship among samples. Based on the sample similarity, we redefine the concept of roughness to measure the uncertainty of the clusters after partitioning. Moreover, by considering the roughness and the sample distribution of the fringe region, we also define the partition validity index to measure the performance of the cluster partitions corresponding to different partition thresholds. Second, based on the sample similarity, we introduce the concepts of intra-class sample similarity and inter-class sample similarity to describe the quantitative relationships of samples within a same cluster and samples in different clusters, respectively. Furthermore, by combining the two kinds of sample similarities, we also define the cluster validity index to measure the performance of clustering corresponding to the number of clusters. Third, we design an automatic three-way clustering method based on the proposed partition validity index and cluster validity index, which aims to obtain the appropriate number of clusters and thresholds of three-way partitioning for achieving a high clustering performance. At last, we implement comparative experiments to verify the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 introduces the preliminary knowledge and related work of three-way decisions theory and three-way clustering. Section 3 gives an automatic threshold selection algorithm based on proposed sample similarity. Section 4 gives an automatic cluster number selection algorithm. Section 5 integrates the threshold selection algorithm and the cluster number selection algorithm to implement an automatic three-way clustering method. Section 6 reports the comparison experimental results. Section 7 concludes this paper.

## 2 Preliminary knowledge and related work

In this section, we will introduce the preliminary knowledge and related work on three-way decisions theory and three-way clustering.

### 2.1 Three-way decisions

In classical two-way decisions theory, one only has two kinds of choices: accept it or reject it. However, in many real-world applications, if existing information is insufficient to support decisions that are explicitly accepted or

rejected, no matter which of the two decisions is chosen, it may cause inaccuracy in the decision. Three-way decisions (3WD) theory [29] is an extension of classical two-way decisions (2WD) theory, in which a deferment decision (or non-commitment decision) is adopted when current information is limited or insufficient to make acceptance decision or rejection decision.

In 3WD theory, a set of samples  $X$  can be divided into three pairwise disjoint regions through three-way decisions, namely the positive region POS ( $X$ ), the boundary region BND ( $X$ ) and the negative region NEG ( $X$ ), which respectively correspond to the acceptance decision, the deferment decision and the rejection decision. A simple representation of 3WD is to construct three-way decisions based rules by incorporating a pair of thresholds. To be specific, for a given sample  $x$  and a subset of samples  $X$ ,  $p(X|x)$  is the conditional probability of  $x$  belongs to  $X$ , if we have a pair of thresholds  $(\alpha, \beta)$  with an assumption of  $0 \leq \beta \leq \alpha \leq 1$ , we can make the following 3WD based rules:

- (P): If  $p(X|x) \geq \alpha$ , decide  $x \in \text{POS}(X)$ ;  
 (B): If  $\beta < p(X|x) < \alpha$ , decide  $x \in \text{BND}(X)$ ;  
 (N): If  $p(X|x) \leq \beta$ , decide  $x \in \text{NEG}(X)$ .

For the (P) rule, if the probability of  $x$  belongs to  $X$  is greater than or equal to  $\alpha$ , we make an acceptance decision to classify  $x$  into the positive region of  $X$ . For the (B) rule, if the probability of  $x$  belongs to  $X$  is greater than  $\beta$  and less than  $\alpha$ , we make a deferment decision to classify  $x$  into the boundary region of  $X$ . For the (N) rule, if the probability of  $x$  belongs to  $X$  is less than or equal to  $\beta$ , we make a rejection decision to classify  $x$  into the negative region of  $X$ .

3WD has been widely concerned by researchers. Many results on 3WD have been reported from the perspectives of theory, model and application. Here we take some representative achievements from the three perspectives to introduce the related work of 3WD. In terms of theory, starting from the 3WD concept proposed by Yao, it has gradually expanded from 3W decisions to 3W + X theories, such as 3W classifications [10, 14, 19, 39, 40], 3W attribute reduction [8, 9, 11, 17], 3W clustering [1, 34, 36], 3W active learning [23], 3W concept learning [13], 3W concept lattices [26, 27], etc. Furthermore, Yao [30] discussed a wide sense of 3WD and proposed a trisecting-acting-outcome (TAO) model of 3WD. In addition, Liu and Liang [22] proposed a function based three-way decisions to generalize the existing models. Li et al. [18] generalized 3WD models based on subset evaluation. Hu [5] studied three-way decision spaces. In terms of model, 3WD was combined with existing models to generate a lot of new models, such as 3WD based Bayesian network [4], neighbourhood three-way decision-theoretic rough set model [15], 3WD based intuitionistic fuzzy decision-theoretic rough sets model [20],

dynamic 3WD models [38], three-way enhanced convolutional neural networks (3W-CNN) [42], etc. In terms of application, 3WD has been successfully applied in sentiment analysis [41, 42], image processing [12], medical decision [28], spam filtering [7], software defect detection [16], etc.

## 2.2 Three-way clustering

As an effective data processing method in machine learning and data mining areas, clustering can effectively divide samples without class information into multiple clusters composed of similar samples, thereby determining the sample distribution in each cluster and improving the efficiency of data post-processing.

The three-way clustering is an improvement of the K-means algorithm. To address the problem of inaccurate decision-making caused by inaccurate information or insufficient data in traditional clustering methods, three-way clustering adds the concepts of positive domain and fringe domain to the clusters in the traditional K-means-based results to further divide the clustering results, which forms the three regions of a cluster as core region (represented by  $Co(C)$  for cluster  $C$ ), fringe region (represented by  $Fr(C)$  for cluster  $C$ ) and trivial region (represented by  $Tr(C)$  for cluster  $C$ ), respectively. If  $x \in Co(C)$ , the object  $x$  belongs to the cluster  $C$  definitely; if  $x \in Fr(C)$ , the object  $x$  might belong to  $C$ ; if  $x \in Tr(C)$ , the object  $x$  does not belong to  $C$  definitely. Moreover, assume  $\pi_K = \{U_1, U_2, \dots, U_K\}$  is the  $K$  clusters of the set of all objects  $U$  based on K-means clustering method, the three regions of all clusters are represented by the followings:

$$\pi_K^T = \{Co(\pi_K), Fr(\pi_K), Tr(\pi_K)\}. \quad (1)$$

According to 3WD theory, if  $x \in Co(\pi_K)$ , then  $x$  is determined to belong to a certain cluster in  $\pi_K$ ; if  $x \in Fr(\pi_K)$ , then  $x$  maybe related to multiple clusters in  $\pi_K$ ; if  $x \in Tr(\pi_K)$ , then  $x$  is determined not to belong to a certain cluster in  $\pi_K$ . Usually, we have  $Tr(\pi_K) = U - Co(\pi_K) - Fr(\pi_K) = \emptyset$  and  $Co(\pi_K) \cap Fr(\pi_K) = \emptyset$ .

In three-way clustering, each cluster can be represented by a pair of core region and fringe region. For the  $K$  cluster partitioning of the sample set, it satisfies the following properties:

- (1)  $\forall U_i \in \pi_K, Co(U_i) \neq \emptyset$ ;
- (2)  $U = \bigcup_{i=1}^K Co(U_i) \cup \bigcup_{i=1}^K Fr(U_i)$ .

The property (1) requires that the core domain of each cluster cannot be empty after division, that is, there is at least one sample in each cluster; the property (2) is used to ensure that the clustering method can implement effective division

for all samples. Thus, the three-way clustering result can be represented by follows:

$$\pi_K = \{(Co(U_1), Fr(U_1)), \dots, (Co(U_K), Fr(U_K))\}. \quad (2)$$

For the fringe region  $Fr(U_i)$ , if  $\bigcup_{i=1}^K Fr(U_i) = \emptyset$ , then  $\pi_K = \{Co(U_1), \dots, Co(U_K)\}$  is a two-way clustering result. In the traditional clustering method, all samples can only be divided into uniquely determined clusters. In the three-way clustering method, if the sample is closely related to multiple clusters at the same time, the sample may belong to multiple clusters at the same time. In this case, it is reasonable to classify the sample into the fringe regions of the closely related clusters.

With the in-depth study of clustering methods, more and more scholars believe that traditional clustering methods are based on hard partitioning, which can not truly reflect the actual relationship between samples and clusters. To address this problem, some approaches such as fuzzy clustering, rough clustering and interval clustering, have been proposed to deal with this kind of uncertain relationship between samples and clusters. These approaches are also called soft clustering or overlapping clustering based on the meaning that a sample can belong to more than one cluster [25, 31]. By using the concept of membership instead of the traditional distance-based methods to measure the degree of membership of each sample with a certain cluster, Friedman and Rubín first proposed the concept of fuzzy clustering [3]. Then, Dunn proposed a widely used fuzzy C-means (FCM) clustering algorithm based on the related concepts of fuzzy clustering [2]. Later, through the extension of the concept of rough set theory, Lingras et al. proposed a rough C-means (RCM) clustering method [21], which considers the different influences of positive and fringe regions of each cluster on calculating corresponding center. Yu introduced 3WD theory into traditional clustering and proposed a framework of three-way cluster analysis [32, 33, 35].

### 3 An automatic threshold selection method based on sample similarity

In the three-way clustering methods, each cluster has a corresponding positive region and fringe region. The thresholds for partitioning the regions of each cluster is the key factor to determine the final clustering performance. Different partition thresholds will generate different regions for the clusters. For three-way decisions theory, it makes the deferment decision for the samples with uncertain information by classifying them into the fringe region. The introduction of such fringe region can deduce the possibility of classifying the sample directly into the wrong

cluster. However, for partitioned clusters, the introduction of the fringe region also brings additional uncertainty for the knowledge representation of the clustering result. If the fringe region contains more samples, the uncertainty of the cluster representation is greater. In view of this, in order to select the optimal thresholds making the clusters have the best partition performance, we define the roughness degree by introducing the sample similarity to measure the quantitative change of the sample distribution in the positive and fringe regions. Moreover, based on the roughness degree, we define a novel partitioning validity index to measure the clustering partitions and propose an automatic threshold selection method for categorical data.

#### 3.1 Partitioning validity index

We first give the definition of sample similarity as follows.

**Definition 1** (*Sample similarity*) [11] Let  $IS = (U, A)$  be an information system, where  $U$  denotes the set of all samples, and  $A$  denotes the set of attributes describing each sample, for any two samples  $x_i \in U$  and  $x_j \in U$ , the similarity of  $x_i$  and  $x_j$  is:

$$S_A(x_i, x_j) = \frac{\sum_{k=1}^{|A|} I(x_{ik}, x_{jk})}{|A|}, \quad (3)$$

where  $|A|$  represents the number of attributes. For categorical data, the function  $I(x_{ik}, x_{jk})$  is computed by:

$$I(x_{ik}, x_{jk}) = \begin{cases} 1 & x_{ik} = x_{jk} \\ 0 & x_{ik} \neq x_{jk} \end{cases}. \quad (4)$$

As defined above, the sample similarity is an indicator of how similar between the samples is. The greater the value, the more similar the two samples.

Through the introduction of sample similarity, the related concepts in 3WD theory can be calculated, which are defined as follows.

**Definition 2** (*Equivalence relation*) [11] Given an information system  $IS = (U, A)$ , the equivalence relation based on a threshold  $\alpha$  is represented by  $E_\alpha$ :

$$E_\alpha = \{(x, y) \in U \times U \mid S_A(x, y) > \alpha\}. \quad (5)$$

Similarly, for sample  $x$ , its equivalence class is computed by:

$$[x]_\alpha = \{y \in U \mid S_A(x, y) > \alpha\}. \quad (6)$$

The equivalence relation  $E_\alpha$  can divide  $U$  into several disjoint sample subsets, denoted by  $\pi_\alpha$ . For a given objective subset  $X \subseteq U$ , we can define its lower and upper approximations based on  $\alpha$  as follows:

$$\begin{aligned} \underline{apr}_\alpha(X) &= \{x \in U | [x]_\alpha \subseteq X\}, \\ \overline{apr}_\alpha(X) &= \{x \in U | [x]_\alpha \cap X \neq \emptyset\}. \end{aligned} \tag{7}$$

Furthermore, we can use the positive region  $POS_\alpha(X)$  and the fringe region  $BND_\alpha(X)$  to describe the objective subset  $X$ :

$$\begin{aligned} POS_\alpha(X) &= \underline{apr}_\alpha(X), \\ BND_\alpha(X) &= \overline{apr}_\alpha(X) - \underline{apr}_\alpha(X). \end{aligned} \tag{8}$$

Usually, the positive region  $POS_\alpha(X)$  contains the samples that belong to  $X$  definitely and the fringe region contains the samples that belong to  $X$  possibly, we can use *roughness* to compute the uncertainty of  $X$  represented by two regions.

**Definition 3 (Roughness)** [24] Given an information system  $IS = (U, A)$ , for objective subset  $X \subseteq U$ ,  $RN_\alpha(X)$  is the roughness of  $X$  based on the threshold  $\alpha$ :

$$RN_\alpha(X) = 1 - \frac{|\underline{apr}_\alpha(X)|}{|\overline{apr}_\alpha(X)|} = \frac{|BND_\alpha(X)|}{|POS_\alpha(X)| + |BND_\alpha(X)|}. \tag{9}$$

Given a threshold, the roughness degree is related to the distribution of samples in the positive region and the fringe region for each cluster, which also leads to a monotonicity property between the roughness and the value of the threshold.

**Property 1** Given an information system  $IS = (U, A)$ ,  $X \subseteq U$  is an objective subset of samples, for any two thresholds  $\alpha$  and  $\beta$ , if  $0 \leq \beta < \alpha \leq 1$ , then  $RN_\beta(X) \geq RN_\alpha(X)$ .

**Proof** For any sample  $x \in U$ , if  $\beta < \alpha$ , according to Eq. (6) and Eq. (7), we have  $[x]_\alpha \subseteq [x]_\beta$ ,  $\underline{apr}_\beta(X) \subseteq \underline{apr}_\alpha(X)$  and  $\overline{apr}_\alpha(X) \subseteq \overline{apr}_\beta(X)$ , thus  $\frac{|\underline{apr}_\alpha(X)|}{|\overline{apr}_\alpha(X)|} \geq \frac{|\underline{apr}_\beta(X)|}{|\overline{apr}_\beta(X)|}$ . Based on Eq. (9), now we have  $RN_\beta(X) \geq RN_\alpha(X)$ .  $\square$

As mentioned above, the introduction of fringe region in three-way clustering is a mixed blessing. On the one hand, it can reduce the classification error by classifying samples with less certainty into the fringe region. On the other hand, it also brings additional uncertainty for knowledge representation. By considering the trade-off relationship between the size of fringe region and the roughness, we define a novel partitioning validity index as follows.

**Definition 4 (Partitioning Validity Index)** Given an information system  $IS = (U, A)$ ,  $X \subseteq U$  is a subset of samples, for a specific threshold  $\alpha$ , the partitioning validity index of  $X$  is defined as:

**Table 1** An information system

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
$x_1$	1	1	1	1	1	1
$x_2$	1	0	1	0	1	1
$x_3$	0	1	1	1	0	0
$x_4$	1	1	1	0	0	1
$x_5$	0	0	1	1	0	1
$x_6$	1	0	1	0	1	1
$x_7$	0	0	0	1	1	0
$x_8$	1	0	1	0	1	1
$x_9$	0	0	1	1	0	1

$$PVI_\alpha(X) = (1 - RN_\alpha(X)) \cdot \frac{|BND_\alpha(X) \cap X|}{|X|}. \tag{10}$$

In this definition,  $\frac{|BND_\alpha(X) \cap X|}{|X|}$  represents the ratio of the number of samples belonging to the objective subset  $X$  and within its fringe region to the number of samples in  $X$ .  $0 \leq PVI_\alpha(X) \leq 1$ .

In the three-way clustering problem, we prefer to obtain a partition with low classification error and put the uncertainty sample in the fringe region to delay decision making, which increases the fringe region while also posing the problem of increasing roughness, so we constrain the fringe region to be too large by minimizing the roughness.

**Example 1** Table 1 is an information system, where  $U = \{x_1, x_2, \dots, x_9\}$  and  $A = \{a_1, a_2, \dots, a_6\}$ . For objective subset  $X = \{x_1, x_2\}$ , if the threshold  $\alpha = 0.8$ , we have  $RN_{0.8}(X) = 3/4$  and  $PVI_{0.8}(X) = 3/8$ . Similarly, if  $\alpha = 0.6$ , we have  $RN_{0.6}(X) = 5/5$  and  $PVI_{0.6}(X) = 0$ . Apparently,  $\alpha = 0.8$  is a better choice for objective subset  $\{x_1, x_2\}$ .

### 3.2 An automatic threshold selection algorithm

For the partitioning validity index, its value tends to be 0 when either the fringe region is maximized or the roughness is minimized, so we choose the value of the largest  $PVI$  as the balance point between the roughness and the fringe region size to show the better comprehensive performance of the clustering result. Thus, we can construct an optimization problem for the threshold selection problem, in which the largest  $PVI_\alpha(X)$  is determined by the optimal threshold  $\alpha$ .

$$\arg \max_\alpha PVI_\alpha(X). \tag{11}$$

Since  $PVI_\alpha(X)$  presents a kind of trade-off relationship between the size of fringe region and the roughness, there does not exist a monotonicity property between  $PVI_\alpha(X)$  and  $\alpha$ . In order to obtain the optimal value of  $\alpha$ , we traverse

all candidate partition thresholds by setting the appropriate step size. Algorithm 1 shows the description of the proposed algorithm.

---

**Algorithm 1:** Threshold selection algorithm

---

**Input:** An information table  $IS = (U, A)$ , objective subset  $X$ , step-size  $\Delta_\alpha = 0.01$ .  
**Output:** The optimal threshold  $\alpha$ .

```

1 for  $x_i \in X$  do
2   for  $x_j \in X$  and  $j > i$  do
3     compute the sample similarity  $S_A(x_i, x_j)$  according to Eq. (3);
4   end
5 end
6 choose the minimal value  $S_{\min}$  and the maximal value  $S_{\max}$  from all
  sample similarities;
7  $\alpha = S_{\min}$ ,  $\alpha_c = S_{\min}$ ,  $MaxPVI = 0$ ;
8 repeat
9   compute  $\overline{apr}_{\alpha_c}(X)$  and  $\overline{apf}_{\alpha_c}(X)$  according to Eq. (7);
10  compute  $POS_{\alpha_c}(X)$  and  $BND_{\alpha_c}(X)$  according to Eq. (8);
11  compute  $RN_{\alpha_c}(X)$  according to Eq. (9);
12  compute  $PVI_{\alpha_c}(X)$  according to Eq. (10);
13  if  $PVI_{\alpha_c}(X) > MaxPVI$  then
14     $MaxPVI = PVI_{\alpha_c}(X)$ ;
15     $\alpha = \alpha_c$ ;
16  end
17   $\alpha_c = \alpha_c + \Delta_\alpha$ ;
18 until  $\alpha_c > S_{\max}$ ;
19 return  $\alpha$ ;

```

---

In the algorithm, the sample similarities of all possible pairs are computed first. The minimal value  $S_{\min}$  and the maximal value  $S_{\max}$  with step-size 0.01 constitute the candidate threshold space. For each candidate threshold, we compute the corresponding positive region and fringe region of objective subset  $X$ , and obtain the current partitioning validity index  $PVI$ . Finally, the threshold with respect to the maximal  $PVI$  is outputted as the optimal threshold. Assume objective subset  $X$  has  $n$  samples, the time complexities of computing all sample similarities and iteratively solving for the optimal threshold are  $O(n^2)$  and  $O(n)$ , thus, the time complexity of Algorithm 1 is  $O(n^2)$ .

#### 4 An automatic cluster number selection method based on sample similarity

In the process of clustering, when the clustering method is determined, the number of clusters will become the key influencing factor of clustering performance. Different number of clusters will also result in completely different clustering results. In general, most existing clustering methods set the number of clusters by expert opinion or experimental cross-validation. In this section, we will give an automatic cluster number selection method based on sample similarity.

In general, all samples can be easily divided into two clusters if there exist large inter-class distance between the two clusters and small intra-class distance in each cluster.

Similarly, we can change the number of clusters by changing the inter-class distance between any two clusters and the intra-class distance among in each cluster, thereby affecting the final clustering performance. To select appropriate number of clusters, we first define the intra-class sample similarity and the inter-class sample similarity concepts that can fully reflect the tightness of distribution between samples in each cluster and the difference of sample distribution between various clusters, respectively. Then, by integrating the intra-class sample similarity and the inter-class similarity, we define a clustering validity index to measure the performance of clusters related to the number of clusters.

#### 4.1 Clustering validity index

We first give the definition of intra-class similarity as follows.

**Definition 5 (Intra-class Similarity for a Cluster)** Given an information system  $IS = (U, A)$ , the intra-class similarity  $IntraS(U_l)$  represents the sample similarity of cluster  $U_l \subset U$  under the set of attributes  $A$ , which is computed as:

$$IntraS(U_l) = \frac{2 \cdot \sum_{i=1}^{|U_l|-1} \sum_{j>i}^{|U_l|} S_A(U_{li}, U_{lj})}{|U_l| \cdot (|U_l| - 1)}, \quad (12)$$

where  $|U_l|$  is the number of samples in cluster  $U_l$ ,  $U_{li}$  and  $U_{lj}$  are the  $i$ -th sample and the  $j$ -th sample in the cluster  $U_l$ , respectively.

Similarly, we can define the intra-class similarity for the partition of  $U$  as follows.

**Definition 6 (Intra-class Similarity for All Clusters)** Given an information system  $IS = (U, A)$ , the intra-class similarity  $IntraS(\pi_K)$  represents the sample similarity of all clusters  $\pi_K = \{U_1, U_2, \dots, U_K\}$ , which is computed as:

$$IntraS(\pi_K) = \frac{\sum_{l=1}^K IntraS(U_l)}{K}, \quad (13)$$

where  $U_l \in \pi_K$  is the  $l$ -th cluster in the partition  $\pi_K$ .

Different from the classification problem, the samples in the clustering problem have no label information. Therefore, it is not enough to only consider the intra-class similarity for each cluster during the clustering process. The inter-class similarity is also a kind of measurement to evaluate the clustering performance, which reflects the degree of differentiation of different clusters.

**Definition 7** (*Inter-class Similarity for Two Clusters*) Given an information system  $IS = (U, A)$ ,  $\pi_K = \{U_1, U_2, \dots, U_K\}$  is the partition of  $U$  with  $K$  clusters. The inter-class similarity  $InterS(U_i, U_j)$  between any two clusters  $U_i$  and  $U_j$  under the set of attributes  $A$  is computed as:

$$InterS(U_i, U_j) = \frac{\sum_{m=1}^{|U_i|} \sum_{n=1}^{|U_j|} S_A(U_{im}, U_{jn})}{|U_i| \cdot |U_j|}, \tag{14}$$

where  $U_{im}$  and  $U_{jn}$  are the  $m$ -th sample in cluster  $U_i$  and the  $n$ -th sample in cluster  $U_j$ , respectively.

Moreover, we can define the inter-class similarity for the partition  $\pi_K$  as follows.

**Definition 8** (*Inter-class similarity for all clusters*) Given an information system  $IS = (U, A)$ , the inter-class similarity  $InterS(\pi_K)$  represents the sample similarity of the partition  $\pi_K = \{U_1, U_2, \dots, U_K\}$ , which is computed as:

$$InterS(\pi_K) = \frac{2 \cdot \sum_{i=1}^{K-1} \sum_{j>i}^K InterS(U_i, U_j)}{K \cdot (K - 1)}. \tag{15}$$

As discussed above, both the intra-class similarity for all clusters and the inter-class similarity for all clusters are two important factors that reflect the clustering performance. Thus, by integrating the two kinds of similarities, we define a novel clustering validity index as follows.

**Definition 9** (*Clustering validity index*) Given an information system  $IS = (U, A)$ ,  $\pi_K = \{U_1, U_2, \dots, U_K\}$  are the  $K$  clusters, the clustering validity index for  $\pi_K$  is computed as:

$$CVI(\pi_K) = IntraS(\pi_K) - InterS(\pi_K). \tag{16}$$

This definition tells us that a good clustering result with a large  $CVI(\pi_K)$  will have a large intra-class sample similarity and a small inter-class sample similarity.

**Example 2** For information system as shown in Table 1, assume all samples are divided into the two clusters  $\pi_2 = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_6, x_7, x_8, x_9\}\}$ , then we have  $IntraS(\pi_2) = 0.5305$  and  $InterS(\pi_2) = 0.6111$ , thus the clustering validity index based on  $\pi_2$  is  $CVI(\pi_2) = -0.0806$ . Considering another situation, assume all samples are divided into the three clusters  $\pi_3 = \{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6, x_7, x_8, x_9\}\}$ , then we have  $IntraS(\pi_3) = 0.5833$  and  $InterS(\pi_3) = 0.5796$ , thus the clustering validity index based on  $\pi_3$  is  $CVI(\pi_3) = 0.0037$ . Apparently,  $\pi_3$  is a better choice based on a larger clustering validity index.

## 4.2 An automatic cluster number selection method

For clustering validity index, the larger the value of  $CVI(\pi_K)$ , the better the performance of the clustering result determined by the current cluster number  $K$ . Therefore, to choose the appropriate number of clusters, we can construct an optimization problem based on clustering validity index, which is represented as follows:

$$\arg \max_K CVI(\pi_K). \tag{17}$$

This optimization problem expects to obtain the optimal number of clusters that maximize the clustering validity index after clustering. Since the distribution of samples (number of samples) in each cluster is not monotonic with the cluster number  $K$ , the clustering validity index does not satisfy the monotonicity property. Thus, we traverse all candidate number of clusters to obtain the optimal  $K$ . Algorithm 2 shows the description of the proposed algorithm. In the algorithm, based on the suggestion in Ref. [37], the search range of  $K$  is set from 2 to  $\sqrt{|U|}$  with step-size 1. Then, we apply K-means clustering algorithm to calculate the initial clustering result and compute the corresponding clustering validity index. Finally, the optimal clustering number with respect to the maximal clustering validity index  $CVI(\pi_K)$  is outputted. Assume  $|U| = n$ , since the time complexity of K-means is  $O(nK)$ , then the time complexity of Algorithm 2 is  $O(n^{\frac{3}{2}}K)$ . Usually,  $K$  is a constant, thus we can say that the final time complexity is  $O(n^{\frac{3}{2}})$ .

---

### Algorithm 2: Cluster number selection algorithm

---

**Input:** An information table  $IS = (U, A)$ , step-size  $\Delta_K = 1$ .  
**Output:** The optimal cluster number  $K$ .  
1 initialize  $K = 2, K^* = 2, MaxCVI = -100$ ;  
2 repeat  
3     apply K-means to obtain the clustering result  $\pi_K$ ;  
4     compute  $CVI(\pi_K)$  for  $\pi_K$  according to Eq. (16);  
5     if  $CVI(\pi_K) > MaxCVI$  then  
6          $MaxCVI = CVI(\pi_K)$ ;  
7          $K^* = K$ ;  
8     end  
9      $K = K + \Delta_K$ ;  
10 until  $K \leq \sqrt{|U|}$ ;  
11 return  $K^*$ ;

---

## 5 An automatic three-way clustering method based on sample similarity

In the fuzzy or rough clustering methods, the influence of the positive region and the fringe region on the cluster centers is different. Similarly, in the three-way clustering method,

the positive region and the fringe region also reflect different influences on the cluster centers. Therefore, based on the threshold  $\alpha$  and the sample similarity, the cluster center in three-way clustering method can be defined as follows.

**Definition 10** (*Cluster center in three-way clustering method*) Given an information system  $IS = (U, A)$ ,  $\pi_K = \{U_1, U_2, \dots, U_K\}$  are  $K$  clusters determined by K-means method. For each cluster  $U_i$ , assume  $\alpha_i$  is the optimal threshold. Let  $PU = POS_{\alpha_i}(U_i)$  and  $BU = BND_{\alpha_i}(U_i)$  represent the positive region and the fringe region of  $U_i$  based on  $\alpha_i$ , respectively.  $x_m$  and  $x_n$  represent the  $m$ -th sample in  $PU$  and the  $n$ -th sample in  $BU$ , respectively. Thus,  $x_m = [x_{m1}, x_{m2}, \dots, x_{m|A|}]$  and  $x_n = [x_{n1}, x_{n2}, \dots, x_{n|A|}]$ . Then the center of cluster  $U_i$  is represented by  $v_i$  which is also a vector with  $|A|$  attribute values.  $v_i = [v_{i1}, v_{i2}, \dots, v_{i|A|}]$ . And the value of  $v_i$  on attribute  $j$  can be computed as:

$$v_{ij} = \frac{\omega_P \cdot \sum_{m=1}^{|PU|} x_{mj} + \omega_B \cdot \sum_{n=1}^{|BU|} x_{nj}}{\omega_P \cdot |PU| + \omega_B \cdot |BU|}, \quad (18)$$

where  $\omega_P$  and  $\omega_B$  are two weight factors for the positive region and the fringe region.

In general,  $\omega_P > \omega_B \geq 0$ . In our experiments,  $\omega_P$  and  $\omega_B$  are set to 1 and 0.5, respectively. From the calculation of the cluster centers and above definitions, it can be seen that the change of the threshold for three-way clustering will also affect the division of the positive and the fringe regions for each cluster, thus affecting the update of cluster centers and the final clustering result. Similarly, the selection of the number of clusters also affects the selection of optimal thresholds and the division of different regions. Therefore, it is not appropriate to separate the selection of the optimal number of clusters and the selection of optimal threshold in the three-way clustering process. In view of this, in this paper, through the integration of the two concepts of clustering validity index and partitioning validity index, an automatic three-way clustering method based on sample similarity is proposed, which can consider the interaction between the change of number of clusters and the variation of the threshold, and results in an optimal three-way clustering result based on the optimal number of clusters and the optimal thresholds. Algorithm 3 shows the description of the proposed algorithm.

**Algorithm 3:** An automatic three-way clustering algorithm based on sample similarity.

---

**Input:** An information table  $IS = (U, A)$ ,  $\Delta_K = 1$ ,  $\epsilon = 0.00001$ .  
**Output:** The positive regions of all clusters  $POSset$  and the fringe regions of all clusters  $BNDset$ .

---

```

1 initialize  $K = 2$ ,  $K^* = 2$ ,  $MaxCVI = -100$ ;
2 randomly select 2 centers for  $U$ :  $\{v_1, v_2\}$ ;
3 repeat
4   compute the distance from each sample in  $U$  to each cluster center
   and merge it with the closest cluster;
5   compute the optimal thresholds  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$  for all clusters
   according to Algorithm 1;
6   compute the corresponding positive and the fringe regions for each
   cluster according to Eq. (8);
7   update the cluster centers according to Eq. (18);
8   compute the distance  $dis$  from the older cluster centers to the new
   cluster centers;
9   if  $dis < \epsilon$  then
10    compute  $CVI(\pi_K)$  for current  $\pi_K$ ;
11    if  $CVI(\pi_K) > MaxCVI$  then
12       $MaxCVI = CVI(\pi_K)$ ;
13       $K^* = K$ ;
14    end
15     $K = K + \Delta_K$ ;
16    randomly select  $K$  centers for  $U$ :  $\{v_1, v_2, \dots, v_K\}$ ;
17  end
18 else
19   break the loop and go to line 22;
20 end
21 until  $K \leq \sqrt{|U|}$ ;
22 apply K-means on  $U$  with  $K^*$  clusters and obtain the clustering result
 $\pi_{K^*} = \{U_1, \dots, U_{K^*}\}$ ;
23 apply Algorithm 1 on  $\pi_{K^*}$  to obtain the optimal thresholds
 $\alpha^* = \{\alpha_1, \dots, \alpha_{K^*}\}$ ;
24 get the final  $POSset$  and  $BNDset$  according to Eq. (8) and  $\alpha^*$ ;
25 return  $POSset$  and  $BNDset$ 

```

---

In this algorithm, the Euclidean distance is used to calculate the distance of any two samples. The optimal threshold based on Algorithm 1 is first selected. When iteratively updating the cluster center changes, if the distance between the older and the new centers is less than  $\epsilon$ , the center is no longer changed and the iteration stops. The front part of the algorithm is mainly to combine Algorithm 1 and Algorithm 2 to obtain the optimal thresholds and the optimal number of clusters, and the latter part is to give the three-way representation of the clustering result. The time complexity of Algorithm 3 is  $O(n^{\frac{5}{2}})$ .

## 6 Experiments

### 6.1 Experiments setting

To evaluate the efficiency of proposed method, we implement several comparison experiments on 15 categorical data sets coming from UCI.<sup>1</sup> The brief descriptions of these data sets are listed in Table 2. In these data sets, some attributes representing ID information have been removed in the preprocessing procedure. Since our data sets are all categorical attributes, it is not easy to calculate

<sup>1</sup> <https://archive.ics.uci.edu/ml/datasets.php>.



**Table 2** The brief descriptions of 15 data sets

ID	Data set	U	A	D
1	balance-scale	625	4	3
2	vote	435	16	2
3	zoo	101	16	7
4	colic	368	22	2
5	heart-statlog	270	13	2
6	primary-tumor	339	17	22
7	bank	600	10	2
8	promoters	106	57	2
9	spect-train	80	22	2
10	hepatitis	155	19	2
11	lung-cancer	32	56	2
12	lymph	148	18	4
13	heart-c	296	13	5
14	house-votes-84	232	16	2
15	breast-cancer	227	9	2

|U| represents the number of samples, |A| represents the number of attributes, and |D| represents the number of classes

the distance of any two samples. Therefore, to facilitate the calculation of the sample distance, we applied one-hot encoding to convert categorical data to integer data.

We compare our method with four kinds of clustering methods including K-means, FCM [2], RCM [21] and a classical three-way clustering method (TWC) [32]. The evaluation measures are the accuracy (ACC) and the normalized mutual information (NMI) that are common used in clustering related works. The clustering accuracy is measured by counting the number of correctly assigned samples and dividing by the number of all samples, which can be computed as follows:

$$ACC = \frac{\sum_{i=1}^{|\pi_L|} |\max\{|\pi_{L_i} \cap \pi_{C_j}| \mid \pi_{C_j} \in \pi_C\}|}{|U|}, \tag{19}$$

where  $\pi_C$  and  $\pi_L$  are real clusters in  $U$  and predicted clusters, respectively.  $\pi_{L_i}$  represents the  $i$ -th cluster of  $\pi_L$  and  $\pi_{C_j}$  represents the  $j$ -th cluster of  $\pi_C$ . The normalized mutual information is used to measure the similarity between two clustering results.

$$NMI(\pi_C, \pi_L) = 2 \cdot \frac{I(\pi_C; \pi_L)}{H(\pi_C) + H(\pi_L)}, \tag{20}$$

where  $I(\pi_C; \pi_L)$  represents the mutual information of  $\pi_C$  and  $\pi_L$ ,  $H(\pi_C)$  and  $H(\pi_L)$  are the related entropy of  $\pi_C$  and  $\pi_L$ , respectively.

**Table 3** The accuracy of all comparison methods on 15 data sets

ID	K-means	FCM	RCM	TWC	Proposed
1	0.6352	0.5312	0.5312	0.6000	<b>0.9216</b>
2	0.8804	0.8804	0.8804	0.8804	<b>0.9793</b>
3	<b>0.6039</b>	<b>0.6039</b>	<b>0.6039</b>	<b>0.6039</b>	0.4059
4	0.6929	0.6847	0.7038	0.7119	<b>0.7527</b>
5	<b>0.8370</b>	0.8296	0.7481	0.8333	0.6666
6	0.2477	0.2536	0.2477	0.2477	<b>0.3215</b>
7	0.5433	0.5550	0.5433	0.5500	<b>0.7016</b>
8	0.6509	<b>0.8301</b>	0.5471	0.6792	0.7358
9	0.6875	0.7000	0.6125	0.6875	<b>0.8125</b>
10	0.7935	0.7935	0.7935	0.7935	<b>0.8709</b>
11	0.7187	0.7187	0.7187	0.7187	<b>0.8437</b>
12	0.7027	0.5472	0.6216	0.6283	<b>0.7297</b>
13	0.8141	0.8243	0.6351	0.8108	<b>0.8817</b>
14	0.8922	0.8879	0.8836	0.8922	<b>1.0000</b>
15	0.7075	0.7075	0.7075	0.7075	<b>0.8122</b>
Score	2	2	1	1	12

The best performance of each row is boldfaced

**Table 4** The normalized mutual information of all comparison methods on 15 data sets

ID	K-means	FCM	RCM	TWC	Proposed
1	0.1016	0.1040	0.1023	0.1016	<b>0.1215</b>
2	0.4525	0.4525	0.4525	0.4525	<b>0.4648</b>
3	0.1089	0.1089	0.1089	0.1089	<b>0.1335</b>
4	0.4389	0.4392	0.4396	0.4396	<b>0.4417</b>
5	<b>0.4423</b>	0.4417	0.4368	0.4420	0.4384
6	0.0306	0.0312	0.0307	0.0307	<b>0.0457</b>
7	0.2811	0.2803	0.2818	0.2805	<b>0.2826</b>
8	0.4225	<b>0.4279</b>	0.4131	0.4194	0.4229
9	0.4206	0.4185	0.4300	0.4185	<b>0.4306</b>
10	0.2738	<b>0.2911</b>	0.2869	0.2755	0.2826
11	0.3918	0.3957	0.3846	0.3846	<b>0.4174</b>
12	0.2058	0.2028	0.2039	0.2040	<b>0.2065</b>
13	0.2794	0.2927	<b>0.2952</b>	0.2801	0.2890
14	0.4466	0.4457	0.4459	0.4466	<b>0.4636</b>
15	0.4344	0.4353	0.4342	0.4344	<b>0.4482</b>
Score	1	2	1	0	11

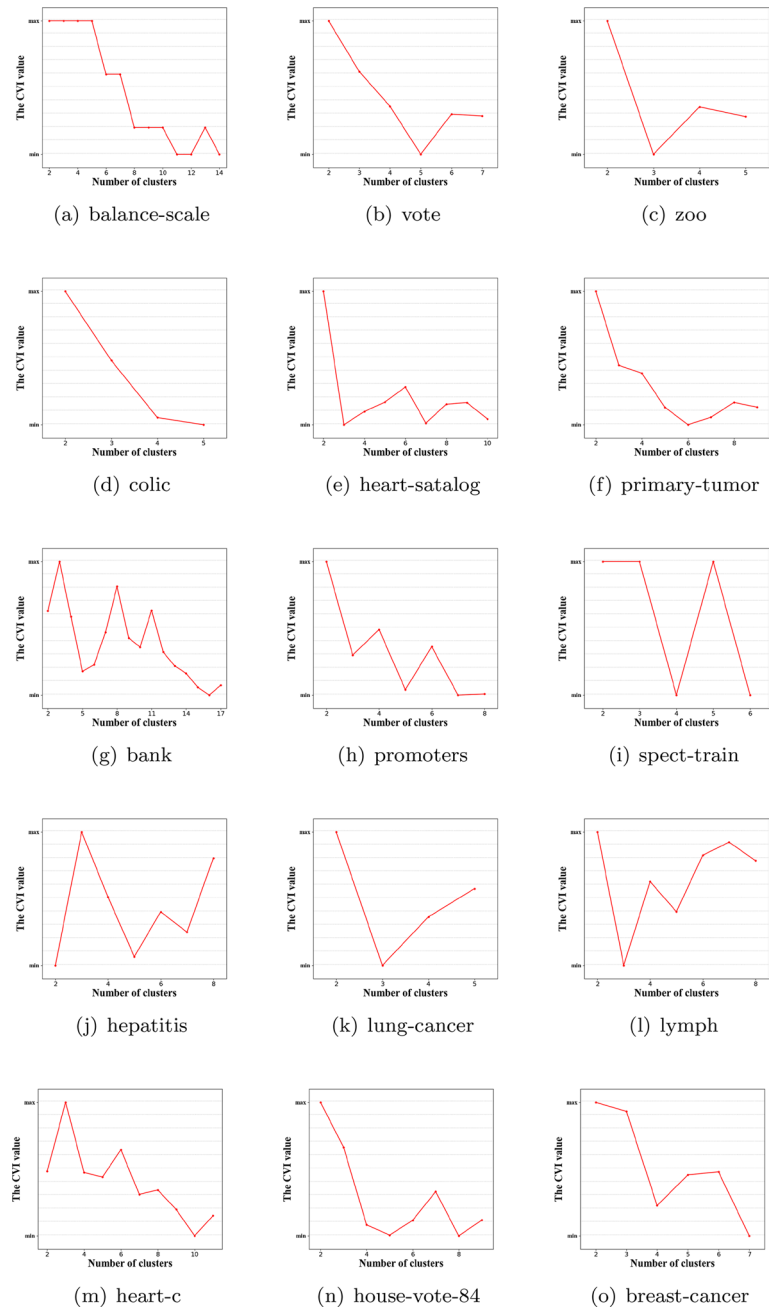
The best performance of each row is boldfaced

### 6.2 Experimental result of clustering performance

The comparison results on ACC and NMI are show in Tables 3 and 4. In each table, ‘score’ means the number of best results in the corresponding clustering method.

From Table 3, we can see that our proposed method can significantly improve the clustering accuracy on 12 data sets.

**Fig. 1** The trend graph of clustering validity index with the change of the number of clusters on each data set



Both our algorithm and the comparison algorithm have low accuracy on datasets with a large number of actual classes, such as data set *zoo* with 7 classes and data set *primary-tumor* with 22 classes, which also indicates that our algorithm performs better on datasets with a small number of actual classes.

Similarly, Table 4 shows that our proposed method can get 11 best results on 15 data sets. Specially, for the NMI measures, our proposed method can beat TWC on all data sets except *heat-stalog* data set.

### 6.3 Experimental result of number of clusters

We also examined the number of clusters based on our proposed clustering method. Table 5 compares our proposed method and the ground-truth result. It can be seen that our proposed method can obtain the true number of clusters on 8 data sets and approximate results on 5 data sets.

Another interesting phenomenon is that our algorithm tends to cluster into very few clusters, such as the datasets *zoo* and *primary-tumor*, which actually have 7 and 22

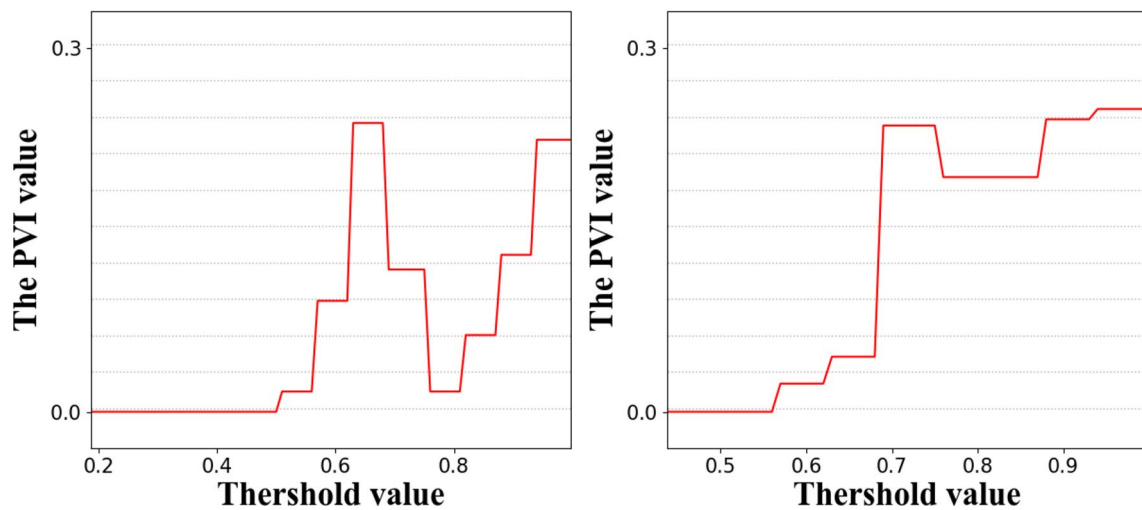


Fig. 2 The trend graph of partitioning validity index with the change of the threshold on data set zoo

Table 5 The comparison between the number of clusters calculated by the proposed algorithm and the actual number of clusters

ID	Proposed	Ground-truth
1	2	3
2	2	2
3	2	7
4	2	2
5	2	2
6	2	22
7	3	2
8	2	2
9	2	2
10	3	2
11	2	2
12	2	4
13	3	5
14	2	2
15	2	2

classes, but our algorithm divides them both into two clusters. On this point we suspect that the clusters themselves are not directly distinguishable, and that many of them are merged into the fringe region of the two clusters. The tendency to use fewer clusters to represent the data and to put the uncertainty into the fringe region is itself a feature of three-way clustering method.

Figure 1 is the trend graph of the clustering validity index with the change of the number of clusters on 15 data sets. In the above comparison experiments, the number of clusters with the largest CVI value of each data set is selected.

### 6.4 Experimental analysis of partitioning threshold

In order to visually show the trend of the partitioning validity index with the partitioning threshold selected by the cluster, we also analyze the PVI values of the two clusters of the data set zoo with the change of the selected partitioning threshold. In Fig. 2, for the two clusters in zoo, the selected thresholds vary by [0.19, 1.0] and [0.44, 1.0], respectively. In the above comparison experiments, we also choose the thresholds with the largest PVI values.

### 7 Conclusion

As a kind of extension of traditional clustering method, three-way clustering method can effectively deal with the incomplete and inaccurate data. In view of the problem that existing three-way clustering methods often select the appropriate number of clusters and the partitioning threshold according to the subjective tuning in the implementation process, in this paper, based on introducing the sample similarity, we defined the clustering validity index and the partitioning validity index to automatically compute the number of clusters and the partitioning threshold, respectively. Furthermore, we proposed an automatic three-way clustering approach by combining the proposed cluster number selection method and the threshold selection method. Comparison experiments also validate the effectiveness of our proposed method.

## References

1. Afridi MK, Azam N, Yao J, Alanazi E (2018) A three-way clustering approach for handling missing data using GTRS. *Int J Approx Reason* 98:11–24
2. Dunn J (1973) A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *J Cybern* 3(3):32–57
3. Friedman HP, Rubin J (1967) On some invariant criteria for grouping data. *J Am Stat Assoc* 62(320):1159–1178
4. Gu Y, Jia X, Shang L (2015) Three-way decisions based bayesian network. In: Proceedings of the IEEE international conference on progress in informatics and computing (PIC), pp 51–55
5. Hu B (2017) Three-way decisions based on semi-three-way decision spaces. *Inf Sci* 382–383:415–440
6. Jain AK, Murty MN, Flynn PJ (1999) *ACM Comput Surv* 31:264–323
7. Jia X, Shang L (2014) Three-way decisions versus two-way decisions on filtering spam email. In: Transactions on rough sets XVIII, pp 69–91
8. Jia X, Liao W, Tang Z, Shang L (2013) Minimum cost attribute reduction in decision-theoretic rough set models. *Inf Sci* 219:151–167
9. Jia X, Shang L, Zhou B, Yao Y (2016) Generalized attribute reduct in rough set theory. *Knowl Based Syst* 91:204–218
10. Jia X, Li W, Shang L (2019) A multiphase cost-sensitive learning method based on the multiclass three-way decision-theoretic rough set model. *Inf Sci* 485:248–262
11. Jia X, Rao Y, Shang L, Li T (2020) Similarity-based attribute reduction in rough set theory: a clustering perspective. *Int J Mach Learn Cybernet* 11:1047–1060
12. Li H, Zhang L, Zhou X, Huang B (2017) Cost-sensitive sequential three-way decision modeling using a deep neural network. *Int J Approx Reason* 85:68–78
13. Li J, Huang C, Qi J, Qian Y, Liu W (2017) Three-way cognitive concept learning via multi-granularity. *Inf Sci* 378:244–263
14. Li W, Huang Z, Jia X (2013) Two-phase classification based on three-way decisions. In: Proceedings of the international conference on rough sets and knowledge technology, pp 338–345
15. Li W, Huang Z, Jia X, Cai X (2016) Neighborhood based decision-theoretic rough set models. *Int J Approx Reason* 69:1–17
16. Li W, Huang Z, Li Q (2016) Three-way decisions based software defect prediction. *Knowl-Based Syst* 91:263–274
17. Li W, Jia X, Wang L, Zhou B (2019) Multi-objective attribute reduction in three-way decision-theoretic rough set model. *Int J Approx Reason* 105:327–341
18. Li X, Yi H, She Y, Sun B (2017) Generalized three-way decision models based on subset evaluation. *Int J Approx Reason* 83:142–159
19. Li Y, Zhang L, Xu Y, Yao Y, Lau RYK, Wu Y (2017) Enhancing binary classification by modeling uncertain boundary in three-way decisions. *IEEE Trans Knowl Data Eng* 29(7):1438–1451
20. Liang D, Xu Z, Liu D (2017) Three-way decisions with intuitionistic fuzzy decision-theoretic rough sets based on point operators. *Inf Sci* 375:183–201
21. Lingras P, Yan R, West C (2003) Comparison of conventional and rough k-means clustering. In: Proceedings of the international conference on rough sets, fuzzy sets, data mining, and granular computing, pp 130–137
22. Liu D, Liang D (2014) An overview of function based three-way decisions. In: Proceedings of the international conference on rough sets and knowledge technology, pp 812–823
23. Min F, Liu F, Wen L, Zhang Z (2019) Tri-partition cost-sensitive active learning through kNN. *Soft Comput* 23:1557–1572
24. Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11(5):341–356
25. Peters G, Crespo F, Lingras P, Weber R (2013) Soft clustering - fuzzy and rough approaches and their extensions and derivatives. *Int J Approx Reason* 54(2):307–322
26. Qi J, Qian T, Wei L (2016) The connections between three-way and classical concept lattices. *Knowl-Based Syst* 91:143–151
27. Qian T, Wei L, Qi J (2017) Constructing three-way concept lattices based on apposition and subposition of formal contexts. *Knowl-Based Syst* 116:39–48
28. Yao J, Azam N (2015) Web-based medical decision support systems for three-way medical decision making with game-theoretic rough sets. *IEEE Trans Fuzzy Syst* 23(1):3–15
29. Yao Y (2010) Three-way decisions with probabilistic rough sets. *Inf Sci* 180:341–353
30. Yao Y (2018) Three-way decision and granular computing. *Int J Approx Reason* 103:107–123
31. Yu H (2018) Three-way decisions and three-way clustering. In: Proceedings of the international joint conference on rough sets, pp 13–28
32. Yu H, Wang Y (2012) Three-way decisions method for overlapping clustering. In: Proceedings of international conference on rough sets and current trends in computing, pp 277–286
33. Yu H, Liu Z, Wang G (2014) An automatic method to determine the number of clusters using decision-theoretic rough set. *Int J Approx Reason* 55(1, Part 2):101–115
34. Yu H, Zhang C, Wang G (2016) A tree-based incremental overlapping clustering method using the three-way decision theory. *Knowl-Based Syst* 91(1):189–203
35. Yu H, Chen Y, Lingras P, Wang G (2019) A three-way cluster ensemble approach for large-scale data. *Int J Approx Reason* 115:32–49
36. Yu H, Wang X, Wang G, Zeng X (2020) An active three-way clustering method via low-rank matrices for multi-view data. *Inf Sci* 507:823–839
37. Yu J, Cheng Q (2002) Search range of optimal cluster number in fuzzy clustering methods. *Sci Chin Ser E Technol Sci* 32:274–280 (in Chinese)
38. Zhang Q, Lv G, Chen Y, Wang G (2018) A dynamic three-way decision model based on the updating of attribute values. *Knowl-Based Syst* 142:71–84
39. Zhang Y, Yao J (2017) Gini objective functions for three-way classifications. *Int J Approx Reason* 81:103–114
40. Zhang Y, Miao D, Zhang Z, Xu J, Luo S (2018) A three-way selective ensemble model for multi-label classification. *Int J Approx Reason* 103:394–413
41. Zhang Y, Miao D, Wang J, Zhang Z (2019) A cost-sensitive three-way combination technique for ensemble learning in sentiment classification. *Int J Approx Reason* 105:85–97
42. Zhang Y, Zhang Z, Miao D, Wang J (2019) Three-way enhanced convolutional neural networks for sentence-level sentiment classification. *Inf Sci* 477:55–64

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.