Pattern Recognition 127 (2022) 108651

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Exploring interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data

Zhong Yuan, Hongmei Chen*, Tianrui Li

School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

A R T I C L E I N F O

Article history:
Received 12 February 2021
Revised 7 March 2022
Accepted 12 March 2022
Available online 15 March 2022

Keywords:
Fuzzy rough set theory
Unsupervised attribute reduction
Complementary entropy
Maximal information
Minimal redundancy
Maximal interactivity
Mixed data

A B S T R A C T

Attribute reduction is one of the important applications in fuzzy rough set theory. However, most attribute reduction methods in fuzzy rough theory mainly focus on removing irrelevant or redundant attributes. There are few reports about the method of considering attribute interaction. For this reason, this paper proposes an interactive attribute reduction method for unlabeled mixed data. First, some uncertainty measures based on fuzzy complementary entropy are further defined. Then, based on the proposed uncertainty measure, the attribute evaluation criteria of maximal information, minimal redundancy, and maximal interactivity are developed respectively. As a result, the evaluation index of the attribute importance is established by using the idea of unsupervised maximal information-minimal redundancy-maximal interactivity. Finally, a corresponding algorithm is designed to select attributes. The experimental results show that the proposed algorithm has better performance.

1. Introduction

Classical rough set theory is an effective tool for handling uncertainty, which has been successfully applied to attribute reduction (also called feature selection) [1], classification [2], clustering [3], etc. Duntch et al. introduced information entropy in rough set theory and proposed three conditional entropy for prediction [4]. Since then, information entropy based on rough set theory has caused extensive research in the fields of data mining, machine learning, and pattern recognition. Some new concepts such as complementary entropy [5], rough entropy [6], combination entropy [7], knowledge granularity [8], and combination granularity [9] have been studied. Some of the above uncertainty measures have been applied to attribute reduction [10–13], and good experimental results have been obtained. However, these studies are all based on the ability to divide knowledge to evaluate the uncertainty of the set, and only apply to nominal attributes.

In order to effectively process numerical or mixed attribute data, Dubois and Prade proposed fuzzy rough sets [14]. Further, information entropy in fuzzy rough set theory is studied. For example, aiming at the importance of fuzzy relation, Yager first introduced the concept of information entropy into fuzzy similarity relation, and then discussed the uncertainty measure of fuzzy information system [15]. Mi et al. gave a new fuzzy entropy and applied it to the partition-based fuzzy rough set for the first time [16]. Since then, many fuzzy uncertainty measures have been proposed [17–20]. Among them, only part is applied to attribute reduction. For example, Hu et al. redefined joint entropy and conditional entropy, and applied them to mixed attribute reduction [17,21]. An et al. combined the minimal Redundancy-Maximal Relevancy (mRMR) algorithm with fuzzy entropy, and proposed an mRMR algorithm based on fuzzy entropy [22]. Considering the deficiencies of [21], Zhang et al. proposed a fuzzy conditional information entropy for feature selection [19]. Wang et al. proposed a measure of uncertainty under generalized fuzzy relations and designed a corresponding attribute reduction algorithm [20]. Based on the concept of gain ratio in decision tree theory, Dai et al. proposed an attribute selection method based on fuzzy gain ratio by using fuzzy mutual information [23]. Lin et al. introduced fuzzy mutual information to evaluate the quality of features in multi-label learning [24]. Dai et al. proposed a feature selection strategy based on fuzzy conditional mutual information to regulate the weight of fuzzy information [25]. Recently, Wang et al. constructed a fuzzy rough self-information-based uncertainty measure, and used it to evaluate the importance of attribute subsets [26]. Using fuzzy neighborhood pessimistic multi-granularity conditional entropy, Sun et al. designed a new feature selection forward search algorithm to improve the classification performance of selected feature subsets [27].

* Corresponding author.
E-mail addresses: yuanzhong2799@foxmail.com (Z. Yuan), hmchen@swjtu.edu.cn (H. Chen), trli@swjtu.edu.cn (T. Li).

https://doi.org/10.1016/j.patcog.2022.108651

However, the above attribute reduction methods are all supervised, and the label information of objects must be known in advance. Therefore, they cannot be applied to attribute reduction of unlabeled data. In addition, as far as we know, there is only a small amount of work using fuzzy rough set theory for unsupervised hybrid attribute reduction [21,28,29]. For example, Hu et al. first used fuzzy information entropy to define the attribute significance for unsupervised feature selection [21]. Ganivada et al. proposed an unsupervised feature selection based on granular neural network using fuzzy rough set model [28]. In [29], Mac ParthaláIn et al. introduced an unsupervised feature selection method based on fuzzy rough sets. However, these methods only consider the correlation or redundancy between attributes. In fact, interactivity is also an important factor in evaluating attributes. The interaction attribute may seem irrelevant, but it may have a stronger ability to distinguish when combined with other attributes [30]. Although the feature selection method considering the interaction has been studied, there are few reports on the method of considering the interaction of attributes in the fuzzy rough set theory.

This paper proposes an interactive attribute reduction method for unlabeled mixed data. First, based on fuzzy complementary entropy, fuzzy complementary conditional entropy, fuzzy mutual information, and fuzzy conditional mutual information are defined respectively. The relationship between several uncertainty measures is discussed. Then, based on fuzzy complementary joint entropy, fuzzy mutual information, and fuzzy conditional mutual information, the evaluation criteria of maximal information, minimal redundancy, and maximal interactivity are respectively constructed to express the importance, redundancy, and interactivity between attributes. Thus, the evaluation index of the attribute importance of unsupervised maximal information-minimal redundancy-maximal interaction is obtained. Finally, we Explore Unsupervised Interactive Attribute Reduction (EUIAR) algorithm for selecting attributes. The experimental results show that the proposed algorithm has better performance than some existing algorithms.

The rest of this paper is organized as follows. In the second section, we introduce the preliminary knowledge on FRS theory. In the third section, some related uncertainty measures are discussed. In the fourth section, we propose an interactive attribute reduction method for unlabeled mixed data and design the corresponding algorithm. Further, an example is adopted to illustrate the method in this paper. Experimental results are given in the fifth section. Finally, the conclusion is given in the sixth section.

2. Preliminaries

Let $U = \{x_1, x_2, \ldots, x_n\}$ be a set of non-empty finite objects. If $\mathcal{A}$ is a mapping from U to [0,1], that is, $\mathcal{A} : U \to [0, 1]$, then it is called $\mathcal{A}$ is a fuzzy set on U. $\forall x \in U$, $\mathcal{A}(x)$ is called the membership function of x. The fuzzy set $\mathcal{A}$ can be expressed as $\mathcal{A} = (\mathcal{A}(x_1), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_n))$ or $\sum_{i=1}^{n} \mathcal{A}(x_i)/x_i$.

2.1. Fuzzy relation

Definition 1. A fuzzy relation $\mathcal{R}$ on U is defined as

$$\mathcal{R} : U \times U \to [0, 1]. \tag{1}$$

$\forall (x, y) \in U \times U$, the membership degree $\mathcal{R}(x, y)$ indicates the degree to which x and y have a relationship $\mathcal{R}$. The set of all fuzzy relations on U is denoted as $\mathcal{F}(U \times U)$. Obviously, the fuzzy relation is a special kind of fuzzy sets.

Suppose $\mathcal{R} \in \mathcal{F}(U \times U)$, $\forall x, y \in U$, if it meets the following conditions

1) Reflexivity: $\mathcal{R}(x, x) = 1$;
2) Symmetry: $\mathcal{R}(x, y) = \mathcal{R}(y, x)$;
3) Transitivity: $\mathcal{R}(x, z) \geq \sup_{y \in U} \min\{\mathcal{R}(x, y), \mathcal{R}(y, z)\}$,

then $\mathcal{R}$ is called a fuzzy equivalence relation on U. Besides, if $\mathcal{R}$ only satisfies 1) and 2), then $\mathcal{R}$ is called a fuzzy similarity relation on U. $\forall \mathcal{R}_1, \mathcal{R}_2 \in \mathcal{F}(U \times U)$, we have

1) $\mathcal{R}_1(x, y) \leq \mathcal{R}_2(x, y) \Rightarrow \mathcal{R}_1 \subseteq \mathcal{R}_2$;
2) $(\mathcal{R}_1 \cap \mathcal{R}_2)(x, y) = \min\{\mathcal{R}_1(x, y), \mathcal{R}_2(x, y)\}$;
3) $(\mathcal{R}_1 \cup \mathcal{R}_2)(x, y) = \max\{\mathcal{R}_1(x, y), \mathcal{R}_2(x, y)\}$.

2.2. Fuzzy rough set

FRS model was first proposed by Dubois and Prade [14], which is defined as follows.

Definition 2. Let $\mathcal{R}$ be a fuzzy equivalence relation on U. $\forall \mathcal{X} \in \mathcal{F}(U)$, the lower approximation $\underline{\mathcal{R}}\mathcal{X}$ and upper approximation $\overline{\mathcal{R}}\mathcal{X}$ of $\mathcal{X}$ are a pair of fuzzy sets on U whose membership functions respectively are

$$\underline{\mathcal{R}}\mathcal{X}(x) = \inf_{y \in U} \max\{1 - \mathcal{R}(x, y), \mathcal{X}(y)\}, \tag{2}$$

$$\overline{\mathcal{R}}\mathcal{X}(x) = \sup_{y \in U} \min\{\mathcal{R}(x, y), \mathcal{X}(y)\}. \tag{3}$$

3. Fuzzy complementary entropy and its related uncertainty measures

In recent years, the theory of information entropy has been continuously used to measure the uncertainty of information systems. A variety of different information entropies have been proposed, such as information entropy [4], rough entropy [6], and complementary entropy [5]. This section mainly defines some related uncertainty measures on the basis of fuzzy complementary entropy, and discusses the relationship between them.

Let $U = \{x_1, x_2, \ldots, x_n\}$ and $C = \{c_1, c_2, \ldots, c_m\}$ be a conditional attribute set. $\forall B \subseteq C$, B can induce a fuzzy similarity relation $\mathcal{R}_B$ on U. It can be denoted by fuzzy relation matrix $M(\mathcal{R}_B) = (r_{ij}^B)_{n \times n}$, where $r_{ij}^B = \mathcal{R}_B(x_i, x_j)$, each row $(r_{i1}^B, r_{i2}^B, \ldots, r_{in}^B)$ denotes a fuzzy set. The fuzzy set induced by $\mathcal{R}_B$ is defined as

$$[x_i]_{\mathcal{R}_B} = \frac{r_{i1}^B}{x_1} + \frac{r_{i2}^B}{x_2} + \ldots + \frac{r_{in}^B}{x_n} = (r_{i1}^B, r_{i2}^B, \ldots, r_{in}^B). \tag{4}$$

Without causing confusion, $\mathcal{R}_B$ can be replaced with B.

Let $B = \{c_{k_1}, c_{k_2}, \ldots, c_{k_h}\} (1 \leq h \leq m) \subseteq C$. Obviously, $[x_i]_B$ is a fuzzy set on B. For the determination of $[x_i]_B$, there are several commonly used methods [31]: 1) Intersection method, 2) Distance method, and 3) Correlation coefficient method. In this paper, the intersection method is used, which is calculated as $[x_i]_B = \bigcap_{l=1}^{h} [x_i]_{c_{k_l}}$. The cardinality of $[x_i]_B$ is defined as $|[x_i]_B| = \sum_{j=1}^{n} r_{ij}^B = \sum_{j=1}^{n} \mathcal{R}_B(x_i, x_j)$. Obviously, $1 \leq |[x_i]_B| \leq n$.

Qian et al. studied fuzzy complementary entropy in FRS theory [18], which is defined as follows.

Definition 3. The fuzzy complementary entropy with respect to B is defined as

$$CE(B) = CE(\mathcal{R}_B) = \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_B|}{|U|}\right). \tag{5}$$

It is easy to get $0 \leq CE(B) \leq 1 - \frac{1}{|U|}$. Among them, $\forall x, y \in U$, $\mathcal{R}_B(x, y) = 1$, i.e., $1 - \frac{|[x_i]_B|}{|U|} = 1 - \frac{|U|}{|U|} = 0$, so $CE(B) = 0$. In this case all object pairs are indistinguishable. Therefore, the granulation space is the coarsest at this time. On the contrary, $\forall x \neq y$, $\mathcal{R}_B(x, y) = 0$, i.e., $1 - \frac{|[x_i]_B|}{|U|} = 1 - \frac{|\{x_i\}|}{|U|}$, so $CE(B) = 1 - \frac{1}{|U|}$. At this time, the granulation space is the smallest. If $\mathcal{R}_B$ degenerates into a crisp equivalence relation, the above fuzzy complementary entropy is the same as the classical complementary entropy mentioned in Liang and Shi [6].

Proposition 1. *If $\mathcal{R}_{c_s} \subseteq \mathcal{R}_{c_t}$, then $CE(\mathcal{R}_{c_s}) \geq CE(\mathcal{R}_{c_t})$.*

Proof. Let $\mathcal{R}_{c_s} \subseteq \mathcal{R}_{c_t}$. So $\forall r_{ij}^{c_s}, r_{ij}^{c_t}$, we have $\forall r_{ij}^{c_s} \leq r_{ij}^{c_t}$. So $|[x_i]_{c_s}| = \sum_{j=1}^{n} r_{ij}^{c_s} \leq \sum_{j=1}^{n} r_{ij}^{c_t} = |[x_i]_{c_t}|$, i.e., $1 - \frac{|[x_i]_{c_s}|}{|U|} \geq 1 - \frac{|[x_i]_{c_t}|}{|U|}$. Therefore, there is $CE(\mathcal{R}_{c_s}) \geq CE(\mathcal{R}_{c_t})$. □

Proposition 2. *If $B_1 \subseteq B_2 \subseteq C$, then $CE(B_1) \leq CE(B_2)$.*

Proof. Let $B_1 \subseteq B_2 \subseteq C$, there is $\mathcal{R}_{B_1} \supseteq \mathcal{R}_{B_2}$. So $\forall r_{ij}^{B_1}, r_{ij}^{B_2}$, there is $\forall r_{ij}^{B_1} \geq r_{ij}^{B_2}$. So there is $|[x_i]_{B_1}| = \sum_{j=1}^{n} r_{ij}^{B_1} \geq \sum_{j=1}^{n} r_{ij}^{B_2} = |[x_i]_{B_2}|$, i.e., $1 - \frac{|[x_i]_{B_1}|}{|U|} \leq 1 - \frac{|[x_i]_{B_2}|}{|U|}$. Therefore, there is $CE(B_1) \leq CE(B_2)$. □

Proposition 2 reflects that the fuzzy complementary entropy changes monotonously with the increase of attributes. It shows that fuzzy complementary entropy can be used as an uncertainty index for fuzzy information systems.

The fuzzy complementary entropy in Definition 3 is derived for one fuzzy relation, and then the fuzzy complementary joint entropy is given for multiple fuzzy relations.

Definition 4 ([18]). $\forall B, E \subseteq C$, the fuzzy complementary joint entropy between B and E is defined as

$$CE(E, B) = CE(\mathcal{R}_{E \cup B}) = \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_E \cap [x_i]_B|}{|U|}\right). \tag{6}$$

Corollary 1. *$\forall B_1, B_2, \ldots, B_m \subseteq C$, the fuzzy complementary joint entropy among $B_1, B_2, \ldots, B_m$ is*

$$CE(B_1, B_2, \ldots, B_m) = CE(\mathcal{R}_{\cup_{k=1}^{m} B_k}) = \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|\cap_{k=1}^{m} [x_i]_{B_k}|}{|U|}\right). \tag{7}$$

The fuzzy complementary joint entropy reflects the uncertainty information under the joint action of multiple attribute subsets.

After an attribute subset B is known, the uncertainty of the attribute subset E can be measured by the fuzzy complementary conditional entropy.

Definition 5. The fuzzy complementary conditional entropy of E on B is defined as

$$CE(E|B) = \frac{1}{|U|} \sum_{i=1}^{n} \left(\frac{|[x_i]_B|}{|U|} - \frac{|[x_i]_E \cap [x_i]_B|}{|U|}\right). \tag{8}$$

Proposition 3. $CE(E|B) = CE(E, B) - CE(B)$.

Proof. According to the above definition, we have

$$CE(E, B) - CE(B) = \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_E \cap [x_i]_B|}{|U|}\right) - \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_B|}{|U|}\right)$$
$$= \frac{1}{|U|} \sum_{i=1}^{n} \left(\frac{|[x_i]_B|}{|U|} - \frac{|[x_i]_E \cap [x_i]_B|}{|U|}\right) = CE(E|B). \quad \square$$

Definition 6. The fuzzy complementary mutual information between B and E is defined as

$$CMI(E; B) = \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_B| + |[x_i]_E| - |[x_i]_B \cap [x_i]_E|}{|U|}\right). \tag{9}$$

Proposition 4. *$\forall B, E \subseteq C$, we have*

1) $CMI(E; B) = CMI(B; E)$;
2) $CMI(E; B) = CE(E) - CE(E|B) = CE(B) - CE(B|E)$;
3) $CMI(E; B) = CE(E) + CE(B) - CE(E, B)$.

Proof. By Definition 6, it is obvious that $CMI(B; E) = CMI(E; B)$, so Formula 1) holds.

$$\begin{aligned} CMI(E; B) &= \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_E| + |[x_i]_B| - |[x_i]_E \cap [x_i]_B|}{|U|}\right) \\ &= \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_E|}{|U|}\right) - \frac{1}{|U|} \sum_{i=1}^{n} \left(\frac{|[x_i]_B|}{|U|} - \frac{|[x_i]_E \cap [x_i]_B|}{|U|}\right) \\ &= CE(E) - CE(E|B). \end{aligned}$$

Similarly, $CMI(E; B) = CE(B) - CE(B|E)$, so Formula 2) holds.

From Proposition 3, there is $CMI(E; B) = CE(E) + CE(B) - CE(E, B)$. Therefore, Formula 3) holds. □

In Proposition 4, the complementary mutual information of E and B is the complementary entropy of E minus the complementary conditional entropy of E with respect to B, so the complementary mutual information of E and B reflects the amount of complementary information contained in E and B, which indicates the degree of correlation between them. In the attribute reduction method in this paper, complementary mutual information is considered as redundancy.

To characterize the interaction of attributes, complementary conditional mutual information is defined as follows.

Definition 7. $\forall B, E, P \subseteq C$, under the condition of known P, the fuzzy complementary conditional mutual information of E and B is defined as

$$\begin{aligned} &CCMI(E; B|P) \\ &= \frac{1}{|U|} \sum_{i=1}^{n} \left(\frac{|[x_i]_E \cap [x_i]_B \cap [x_i]_P| + |[x_i]_P| - |[x_i]_E \cap [x_i]_P| - |[x_i]_B \cap [x_i]_P|}{|U|}\right). \end{aligned} \tag{10}$$

Proposition 5. *$\forall B, E, P \subseteq C$, we have*

1) $CCMI(E; B|P) = CCMI(B; E|P)$;
2) $CCMI(E; B|P) = CE(E, P) + CE(B, P) - CE(E, B, P) - CE(P)$.

Proof. From Definition 7, it is obvious that $CCMI(E; B|P) = CCMI(B; E|P)$, so the Formula 1) holds. Since

$$\begin{aligned} &CE(E, P) + CE(B, P) - CE(E, B, P) - CE(P) \\ &= \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_E \cap [x_i]_P|}{|U|}\right) + \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_B \cap [x_i]_P|}{|U|}\right) \\ &\quad - \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_E \cap [x_i]_B \cap [x_i]_P|}{|U|}\right) - \frac{1}{|U|} \sum_{i=1}^{n} \left(1 - \frac{|[x_i]_P|}{|U|}\right) \\ &= \frac{1}{|U|} \sum_{i=1}^{n} \left[\left(1 - \frac{|[x_i]_E \cap [x_i]_P|}{|U|}\right) + \left(1 - \frac{|[x_i]_B \cap [x_i]_P|}{|U|}\right)\right. \\ &\quad \left. - \left(1 - \frac{|[x_i]_E \cap [x_i]_B \cap [x_i]_P|}{|U|}\right) - \left(1 - \frac{|[x_i]_P|}{|U|}\right)\right] \\ &= \frac{1}{|U|} \sum_{i=1}^{n} \left(\frac{\begin{array}{l} |[x_i]_E \cap [x_i]_B \cap [x_i]_P| + |[x_i]_P| - |[x_i]_E \cap [x_i]_P| \\ -|[x_i]_B \cap [x_i]_P| \end{array}}{|U|}\right) \\ &= CCMI(E; B|P). \end{aligned}$$

Therefore, Formula 2) holds. □

The conditional complementary mutual information reflects the amount of complementary information contained in E and B under the condition of known P, which can be used to express the interaction of attributes.

4. Exploring unsupervised interactive attribute reduction

Based on the above fuzzy complementary uncertainty measures, this section explores an unsupervised interactive attribute reduction, which mainly involves method, algorithm, and example.

4.1. Method

In order to effectively process nominal, numerical, and mixed feature data, the fuzzy similarity degree $r_{ij}^{c_k}$ between x_i and x_j with respect to c_k is calculated as Yuan et al. [32]

$$r_{ij}^{c_k} = \begin{cases} 1, f(x_i, c_k) = f(x_j, c_k) \text{ and } c_k \text{ is nominal;} \\ 0, f(x_i, c_k) \neq f(x_j, c_k) \text{ and } c_k \text{ is nominal;} \\ 1 - \left|f(x_i, c_k) - f(x_j, c_k)\right|, |f(x_i, c_k) - f(x_j, c_k)| \leq \varepsilon_{c_k} \text{ and } c_k \text{ is numerical;} \\ 0, |f(x_i, c_k) - f(x_j, c_k)| > \varepsilon_{c_k} \text{ and } c_k \text{ numerical,} \end{cases} \tag{11}$$

where ε_{c_k} is the adaptive fuzzy radius. ε_{c_k} is calculated as $\varepsilon_{c_k} = \frac{std(c_k)}{\lambda}$. $std(c_k)$ is the standard deviation of the attribute value of c_k, and the default parameter λ is used to adjust the fuzzy radius.

The goal of unsupervised attribute reduction is to find an attribute subset that contains most or all of the information contained in the original attribute set. The attribute subset can maximize the preservation of the original attribute information. According to Definition 3 and Proposition 1, fuzzy complementary entropy can be used to measure the amount of information contained in an attribute. The larger the fuzzy complementary entropy of an attribute, the more information it provides for unsupervised learning. If the attribute with the largest fuzzy complementary entropy is selected, the data can lose information to a minimal. Therefore, the unsupervised attribute reduction method in this paper starts with an empty set *Red*, and then selects one attribute at a time in a stepwise manner. In Step 1, the first important attribute c_{k_1} is selected, which should meet the following conditions:

$$c_{k_1} = arg\max_{c_k \in C}\{CE(\{c_k\})\}. \tag{12}$$

Since the attribute c_{k_1} has the largest fuzzy complementary entropy, it can provide more information for unsupervised learning. In other words, when only one attribute is selected, c_{k_1} provides the most information for the fuzzy information system. At this time, let $Red = \{c_{k_1}\}$.

Assuming that C_u represents the set of attributes that are not currently selected, and *Red* represents the reduced set of selected $(r-1)$ attributes, how to select the rth attribute? In the method proposed in this paper, the following method is used for the selection of the rth attribute: the selected attribute c_{k_r} should have the maximal joint information with *Red* and have the minimal redundancy with the selected attribute in *Red*. In addition, it should also have the maximal interaction with other candidate attributes and *Red*.

Fuzzy complementary joint entropy can be used to measure the joint information between two attribute subsets. The larger the value of joint information between two attribute subsets, the greater the correlation between them. Therefore, the selected attribute should have the largest complementary joint entropy with the selected attribute *Red*. Based on the complementary joint entropy, $\forall c \in C_u$, the maximal joint information is formalized as follows.

$$\max_{c \in C_u}\{CE(Red, \{c\})\}. \tag{13}$$

However, if only the maximal joint information is considered, the redundancy between two attributes may be ignored. The attributes selected in this way may be highly dependent on each other and cannot provide additional information. Therefore, if some of the information is eliminated, the unsupervised learning result will not be affected. Fuzzy complementary mutual information can be regarded as the amount of information shared by two attribute subsets, and it can be used for measuring redundancy between two attributes. Therefore, the minimal redundancy is further considered on the basis of the complementary mutual information, which is expressed as follows.

$$\min_{c \in C_u}\left\{\frac{1}{|Red|}\sum_{s=1}^{r-1} CMI(\{c\}; \{c_{k_s}\})\right\}. \tag{14}$$

Therefore, when selecting the rth attribute, if the maximal joint information of the candidate attribute with *Red* and its redundancy with the selected attributes are comprehensively considered, an unsupervised maximal information-minimal redundancy-based the evaluation index can be constructed as follows.

$$c_{k_r} = arg\max_{c \in C_u}\left\{CE(Red, \{c\}) - \frac{1}{|Red|}\sum_{s=1}^{r-1} CMI(\{c\}; \{c_{k_s}\})\right\}. \tag{15}$$

The above evaluation index selects the rth attribute with the least redundancy with the selected attribute while retaining the maximal joint information. However, there is still interaction between attributes. If the interaction between them is ignored, it may be difficult to obtain the optimal result. Therefore, we further explore the interaction of attributes to find a more suitable evaluation index. For this reason, the maximal interactivity based on complementary condition mutual information is defined as

$$\max_{c \in C_u}\left\{\frac{1}{|C_u - \{c\}|}\sum_{c' \in C_u - \{c\}} CCMI(\{c'\}; Red|\{c\})\right\}. \tag{16}$$

Therefore, when selecting the rth important attribute, if the maximal interactivity among the candidate attributes, other candidate attributes, and the selected attribute set *Red* is further considered, then the unsupervised maximal information–minimal redundancy–maximal interactivity-based (UMiMRMI) evaluation index can be constructed as follows.

$$c_{k_r} = arg\max_{c \in C_u}\left\{CE(Red, \{c\}) - \frac{1}{|Red|}\sum_{s=1}^{r-1} CMI(\{c\}; \{c_{k_s}\}) + \frac{1}{|C_u - c|}\sum_{c' \in C_u - c} CCMI(\{c'\}; Red|\{c\})\right\}. \tag{17}$$

Algorithm 1: The algorithm EUIAR.

Input: A data with $|U| = n, |C| = m$, threshold λ, and the number of selected attributes Att_{num}
Output: A reduction Red

1 $Red \leftarrow \emptyset, C_u \leftarrow C$;
2 **for** $k \leftarrow 1$ *to* m **do**
3   Compute $M_{\mathcal{R}_{c_k}}$;
4   Compute $CE(\{c_k\})$;
5 **end**
6 Select the attribute c_{k_1} so that $CE(\{c_{k_1}\})$ has a maximal value ;
7 $Red \leftarrow Red \cup \{c_{k_1}\}, C_u \leftarrow C_u - \{c_{k_1}\}$;
8 **while** $|Red| \le Att_{num}$ **do**
9   **for** $l \leftarrow 1$ *to* $|C_u|$ **do**
10   Calculate $CE(Red, \{c_l\})$;
11   **for** $s \leftarrow 1$ *to* $|Red|$ **do**
12   Calculate $CMI(\{c_l\}; \{c_{k_s}\})$;
13   **end**
14   **for** $c' \in C_u - \{c_{k_r}\}$ **do**
15   Calculate $CCMI(\{c'\}; Red|\{c_l\})$;
16   **end**
17   **end**
18   Select the attribute c_{k_r} so that $CE(Red, c_{k_r}) - \frac{1}{|Red|}\sum_{s=1}^{|Red|} CMI(\{c_{k_r}\}; \{c_{k_s}\}) + \frac{1}{|C_u - \{c_{k_r}\}|}\sum_{s=1}^{|C_u - \{c_{k_r}\}|} CCMI(\{c'\}; Red|\{c_{k_r}\})$ has a maximal value ;
19     $Red \leftarrow Red \cup \{c_{k_r}\}, C_u \leftarrow C_u - \{c_{k_r}\}$;
20 **end**
21 **return** Red.

Therefore, the rth attribute can be selected as c_{k_r}, which can provide the largest amount of information to unsupervised learning. It has the greatest joint information and interactivity, and contains only a few redundant. When selecting subsequent attributes, we use a similar method to select one by one. Finally, a subset of attributes containing Att_{num} attributes is obtained.

4.2. Algorithm

In this section, we design the algorithm EUIAR and analyze its complexity.

Algorithm 1 takes the empty set as the starting point, the fuzzy complementary entropy of each attribute is first calculated, and then the attribute with the largest fuzzy complementary entropy is selected to add to Red. Then, the UMiMRMI evaluation index is used to select subsequent attributes until a subset containing Att_{num} attributes is obtained.

In Algorithm 1, the number of cycles of Steps 2–5 is m, the number of cycles of Step 3 is $n \times n$, and the number of cycles of steps 9–17 is $|C_u| \times |Red|$. Therefore, the total number of cycles of Algorithm 1 is $m \times n \times n + |C_u| \times |Red|$. Therefore, in the worst case, the time complexity of Algorithm 1 is $O(mn^2)$.

4.3. Example

In this subsection, an example is given to illustrate the proposed algorithm.

Example 1. A data table is on the left of Table 1, which mainly involves mixed attribute data. Here, $U = \{x_1, x_2, \dots, x_6\}$, $C = \{c_1, c_2, c_3, c_4\}$. Among them, c_1, c_2 are numeric attributes, and c_3, c_4 are nominal attributes.

The min-max standardization is performed on the original numerical data, and the standardized results are shown on the right side of Table 1. The standard deviations on numerical attributes c_1, c_2 are calculated as $std(c_3) \approx 0.3492$ and $std(c_4) \approx 0.3146$, respectively. According to Formula (11) and $\lambda = 1$, the fuzzy radii are calculated as $\varepsilon_{c_1} \approx 0.3168$ and $\varepsilon_{c_2} \approx 0.3416$, respectively.

Table 1
Initial and standardized fuzzy information systems.

U	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
x_1	8	0.6	f	D	1	0.8	f	D
x_2	4	0.4	g	A	0.4286	0.4	g	A
x_3	6	0.5	g	B	0.7143	0.6	g	B
x_4	3	0.2	e	B	0.2857	0	e	B
x_5	1	0.3	e	C	0	0.2	e	C
x_6	5	0.7	g	A	0.5714	1	g	A

Table 2
The complementary joint entropy of single attribute.

$CE(c_k, c_s)$	c_1	c_2	c_3	c_4
c_1	0.5317	0.7143	0.6587	0.7857
c_2	0.7143	0.6111	0.7444	0.8333
c_3	0.6587	0.7444	0.6111	0.7778
c_4	0.7857	0.8333	0.7778	0.7222

Table 3
The complementary mutual information of attribute.

$CMI(c_k; c_s)$	c_1	c_2	c_3	c_4
c_1	0.5317	0.4286	0.4841	0.4683
c_2	0.4286	0.6111	0.4778	0.5000
c_3	0.4841	0.4778	0.6111	0.5556
c_4	0.4683	0.5000	0.5556	0.7222

For $\forall c_j \in C$, the fuzzy similarity relation matrix is as follows. For example, for attribute c_1, since $|f(x_1, c_1) - f(x_3, c_1)| = |1 - 0.7143| = 0.2857 \le \varepsilon_{c_1}$, so $r_{13}^{c_1} = 1 - |f(x_1, c_1) - f(x_3, c_1)| = 1 - 0.2857 = 0.7143$.

$$M(\mathcal{R}_{c_1}) = \begin{pmatrix} 1 & 0 & 0.7143 & 0 & 0 & 0 \\ 0 & 1 & 0.7143 & 0.8571 & 0 & 0.8571 \\ 0.7143 & 0.7143 & 1 & 0 & 0 & 0.8571 \\ 0 & 0.8571 & 0 & 1 & 0.7143 & 0.7143 \\ 0 & 0 & 0 & 0.7143 & 1 & 0 \\ 0 & 0.8571 & 0.8571 & 0.7143 & 0 & 1 \end{pmatrix}$$

$$M(\mathcal{R}_{c_2}) = \begin{pmatrix} 1 & 0 & 0.8 & 0 & 0 & 0.8 \\ 0 & 1 & 0.8 & 0 & 0.8 & 0 \\ 0.8 & 0.8 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.8 & 0 \\ 0 & 0.8 & 0 & 0.8 & 1 & 0 \\ 0.8 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$M(\mathcal{R}_{c_3}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}, M(\mathcal{R}_{c_4}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

According to Definition 3, the fuzzy complementary entropy is calculated as follows.

$$\begin{aligned} CE(\{c_1\}) = \frac{1}{|U|}\sum_{i=1}^{n}\left(1 - \frac{|[x_i]_{c_1}|}{|U|}\right) &= \frac{1}{6}\left(1 - \frac{1.7143}{6}\right) \\ &+ \frac{1}{6}\left(1 - \frac{3.4285}{6}\right) + \frac{1}{6}\left(1 - \frac{3.2857}{6}\right) + \frac{1}{6}\left(1 - \frac{3.2857}{6}\right) \\ &+ \frac{1}{6}\left(1 - \frac{1.7143}{6}\right) + \frac{1}{6}\left(1 - \frac{3.4285}{6}\right) \approx 0.5317. \end{aligned}$$

Similarly, there are $CE(\{c_2\}) \approx 0.6111, CE(\{c_3\}) \approx 0.6111$ and $CE(\{c_4\}) \approx 0.7222$.

Next, the attribute with the largest fuzzy complementary entropy is selected and added to *Red*, that is, attribute c_4 is added to *Red*, and $Red = \{c_4\}, C_u = \{c_1, c_2, c_3\}$.

Through Formula (6), the fuzzy complementary joint entropy of single attribute is calculated as shown in Table 2.

According to Proposition 3, the complementary mutual information of single attributes is calculated as in Table 3

By Proposition 3, the condition of $\forall c \in C_u$ is added to *Red*, and the conditional complementary mutual information of *Red* and other candidate attributes is calculated as follows.

$CCMI(c_2; Red|c_1) = CE(c_2, c_1) + CE(Red, c_1) - CE(c_2, Red, c_1) - CE(c_1) \approx 0.1349, CCMI(c_3; Red|c_1) \approx 0.1270$;

$\mathrm{CCMI}(c_1; Red|c_2) \approx 0.1032, CCMI(c_3; Red|c_2) \approx 0.1333$;

$\mathrm{CCMI}(c_1; Red|c_3) \approx 0.0397, CCMI(c_2; Red|c_3) \approx 0.0778$;

From this, we can get

$$CE(Red, c_1) - CMI(c_1, c_4) + \frac{1}{2}(CCMI(c_2; Red|c_1) + CCMI(c_3; Red|c_1)) = 0.4484;$$

$$CE(Red, c_2) - CMI(c_2, c_4) + \frac{1}{2}(CCMI(c_1; Red|c_2) + CCMI(c_3; Red|c_2)) = 0.4516;$$

$$CE(Red, c_3) - CMI(c_3, c_4) + \frac{1}{2}(CCMI(c_1; Red|c_3) + CCMI(c_2; Red|c_3)) = 0.2810.$$

Next, c_2 is selected to add to *Red*, we have $Red = \{c_4, c_2\}, C_u = \{c_1, c_3\}$.

Let $Att_{num} = 3$, we use similar strategies to select one by one. Finally, a subset of attribute $Red = \{c_4, c_2, c_1\}$ is obtained.

.

Table 4
The description of dataset.

No.	Dataset		Number of objects	Number of conditional attributes		Decision class
	Original name	Abbr.		Numerical	Nominal	
1	audiology	audi	226	0	69	24
2	balance-scale	bala	625	0	4	3
3	breast-cancer	brea	286	0	9	2
4	monks	monk	432	0	6	2
5	mushroom	mush	8124	0	22	2
6	soybean-small	soyb	47	0	35	4
7	spect	spect	267	0	22	2
8	tic_tac_toe	tic	958	0	9	2
9	vote	vote	435	0	16	2
10	dermatology	derm	366	34	0	6
11	diabetes	diab	768	8	0	2
12	ionosphere	iono	351	33	0	2
13	iris	iris	150	4	0	3
14	movement_libras	move	360	90	0	15
15	waveform	wave	5000	21	0	3
16	wine	wine1	178	13	0	3
17	winequality_white	wine2	4898	11	0	7
18	yeast	yeast	1484	8	0	10
19	abalone	abalone	4177	7	1	28
20	autos	autos	205	15	10	6
21	bands	bands	531	24	15	2
22	flag	flag	194	10	18	8
23	heart	heart	270	6	7	2
24	hepatitis	hepa	155	6	13	2
25	horse	horse	368	7	20	2
26	labor	labor	57	8	8	2
27	sick	sick	3772	6	23	2

5. Experiments

In this section, data experiments on attribute reduction are performed to verify the superiority and effectiveness of the proposed EUIAR. Specific experiments are divided into clustering experiments and parameter sensitivity analysis. Before analyzing the above experimental results, some detailed experimental preparations are introduced first

5.1. Data sets

The proposed method and other comparison methods are evaluated through 27 public data sets. These data sets are downloaded from the UCI database.[1] For some missing values in some data sets, the maximal probability value method is used to fill in the missing attribute values. Finally, the number of objects, the number of attributes, and the actual decision class are shown in Table 4. From Table 4, it can be seen that the experimental data contains three types, namely numerical, nominal, and mixed attribute data.

5.2. Comparison algorithm

On data sets listed in Table 4, seven unsupervised attribute reduction algorithms are used to compare the proposed EUIAR through the clustering task. These comparison algorithms are described as follows.

1) Baseline: The clustering results on all original attribute sets are used as the baseline.
2) UnSupervised Quick Reduct (USQR) [33]: USQR algorithm is an unsupervised quick reduct based on the classic rough positive domain. It usually outputs a reduced subset.
3) Unsupervised Entropy Based Reduct (UEBR) [34]: This algorithm uses conditional entropy to define the importance of attributes, and then outputs a feature subset.
4) Unsupervised Fuzzy Rough-based Feature Selection (UFRFS) [29]: This algorithm uses fuzzy positive domain to define the importance of attributes, and then outputs a feature subset.
5) Fuzzy Entropy-based Unsupervised Attribute Recuction (FEUAR) [21]: The fuzzy entropy is used to construct the importance of attributes in FEUAR, thereby performing unsupervised attribute reduction. Finally, a reduced subset is also obtained.
6) Laplacian Score-based (LS) [35]: LS calculates the importance of features by the consistency of the feature and the Laplacian matrix, and then a feature sequence can be obtained.
7) Feature Similarity based Feature Selection (FSFS) [36]: FSFS constructs the evaluation indicator of feature score through feature similarity to perform feature selection. It outputs a feature sequence arranged in descending order according to the feature score.
8) SPECtral analysis-based (SPEC) [37]: SPEC uses the non-trivial feature vector of the standardized Laplacian matrix to calculate the feature score. Finally, it also outputs a feature sequence.

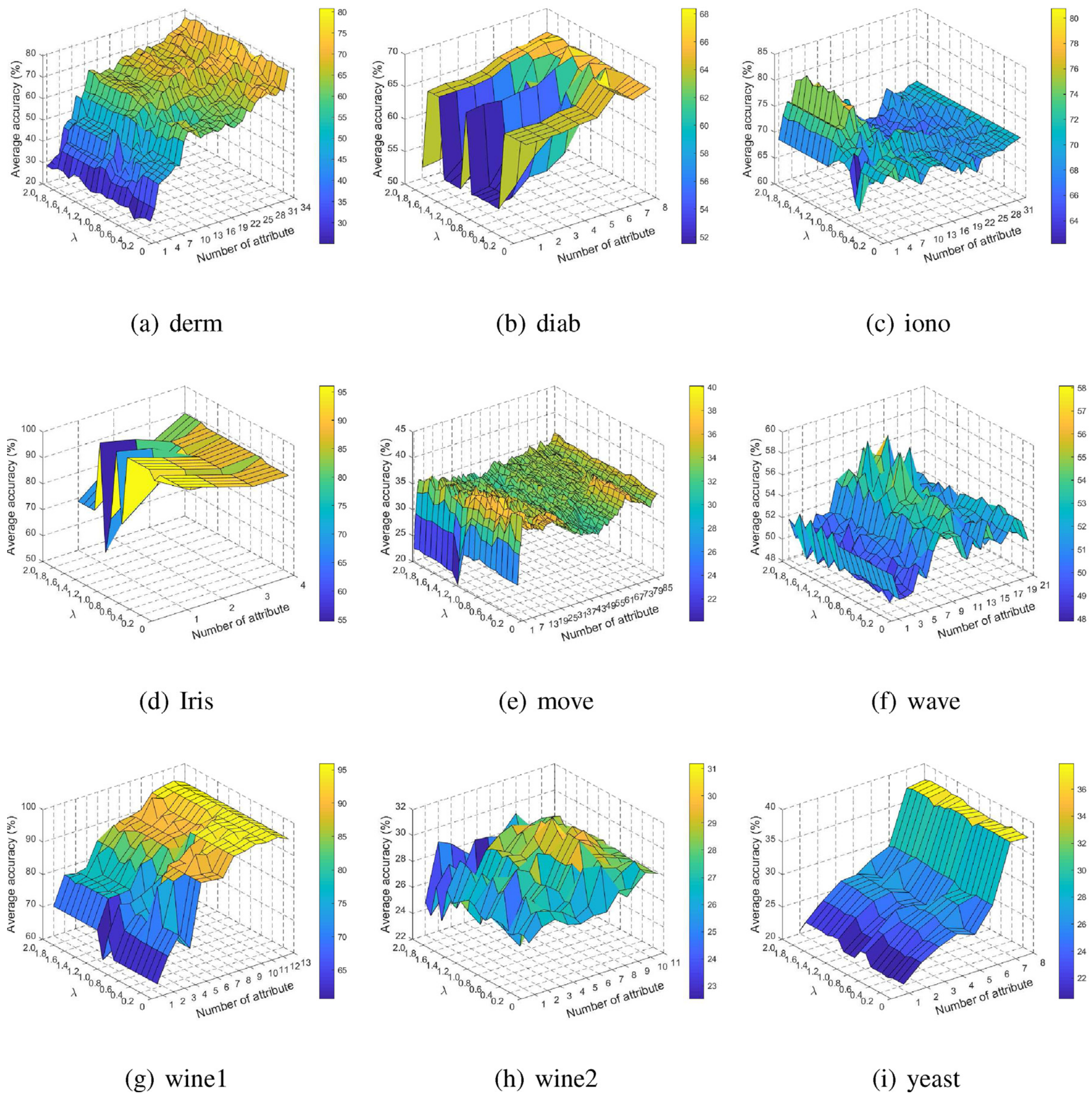
Among them, LS, FSFS and SPEC are usually applied to numerical attribute data. Both USQR and UEBR are attribute reduction algorithms based on rough sets, and they are only applicable to nominal attribute data. Therefore, the numerical attribute data needs to be discretized preprocessing. UFRFS and FEUAR are methods based on fuzzy rough set theory, which can be conducted on both numerical and categorical data.

5.3. Experimental settings

Clustering algorithms k-Means, k-Medoids, and Fuzzy C-Means (FCM) are employed to evaluate the performance of the above feature selection methods. Usually, the clustering algorithm returns the predicted labels of a data set. The clustering Accuracy (Acc) can be used to evaluate the performance of clustering methods [38], which is defined as follows.

$$Acc = \frac{\sum_{i=1}^{n} \delta(f(x_i, d), f(x_i, d'))}{n}, \tag{18}$$

[1] http://archive.ics.uci.edu/ml



(a) derm (b) diab (c) iono

(d) Iris (e) move (f) wave

(g) wine1 (h) wine2 (i) yeast

Fig. 1. The average clustering accuracy varies with λ and the number of attributes on numerical data sets

where $f(x_i, d)$ and $f(x_i, d')$ represent the true decision attribute value and predicted decision attribute value of x_i respectively. If $f(x_i, d) = f(x_i, d')$, then $\delta(f(x_i, d), f(x_i, d')) = 1$. Otherwise, it is equal to 0. The larger the Acc of a clustering algorithm, the better its performance. Because of the randomness of these clustering algorithms, we repeat the experiment 10 times. Finally, the average and standard deviation of the clustering accuracy are recorded as the experimental result.

In the experiment, for LS, FSFS, and SPEC, all different nominal attribute values are replaced with different integer values, and all attribute values are normalized to the interval [0,1] using min-max normalization. Following the preferred experimental setting in Zhu et al. [38], the neighborhood size is fixed to be 5 when LS calculates the neighbor graph matrix. For FSFS, the feature similarity is calculated by "Maximal Information Compression Index". For FSFS, its "style" is set as -1. Usually, LS, FSFS, and SPEC output a feature sequence, and the preferred number of selected features is unknown. Therefore, we calculate the best result with the number of features in the range of $[1, m-1]$. Since USQR and UEBR only consider nominal attributes, the discretization method needs to be used to preprocess numerical attributes. The experimental results in Hu et al. [21] show that the FCM discretization method provides a better performance than equal width and equal frequency. Therefore, FCM discretization is employed to discretize numerical data. Among them, the numerical attributes are discretized into 4 intervals. For EUIAR, we calculate the optimal feature subset of λ in the range of [0.1,2] with step size 0.1.

5.4. Clustering results

In this subsection, the selected attributes and accuracy of clustering algorithms are presented. Further, the analyses on these results are also given.

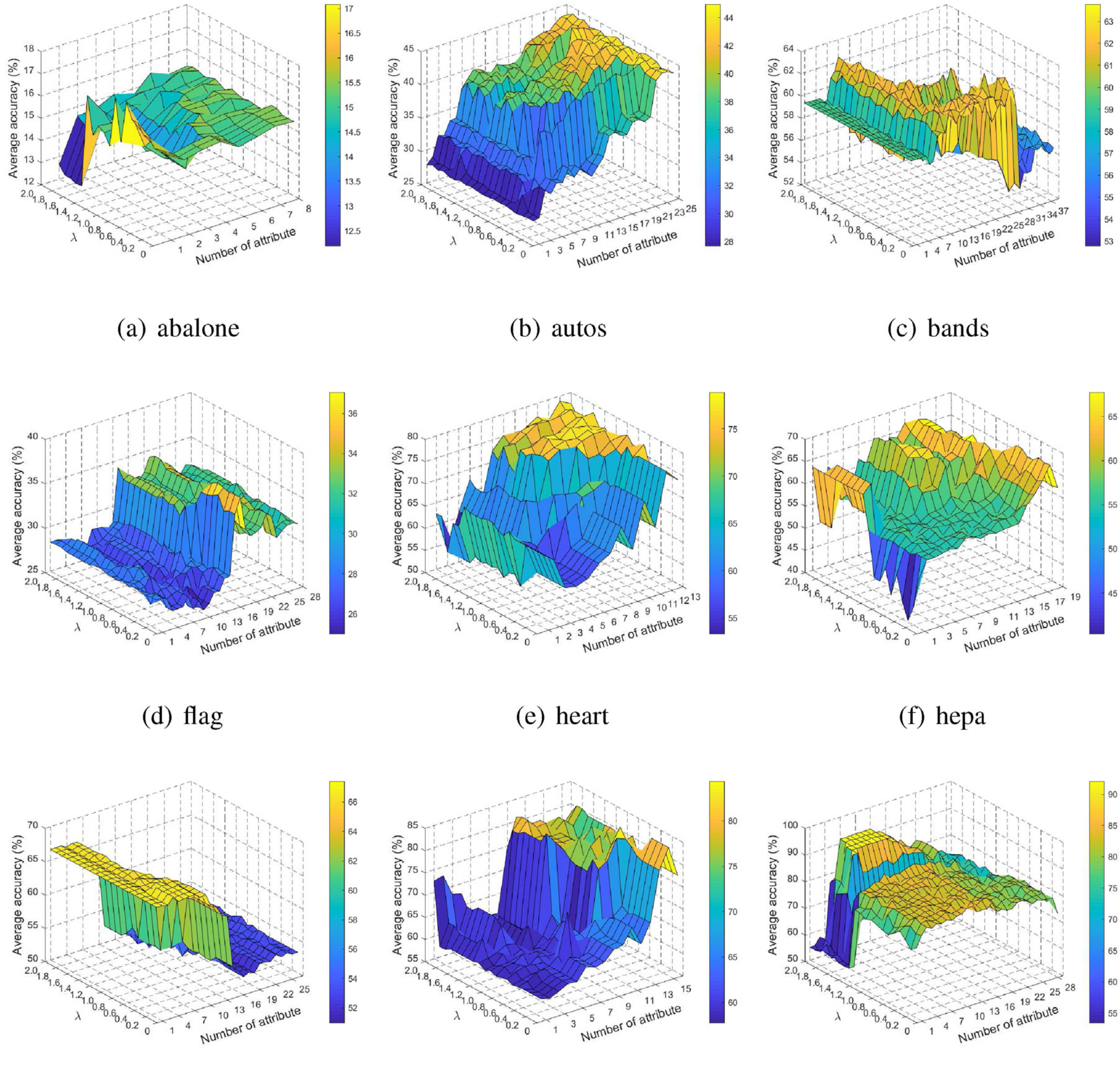
5.4.1. Selected attributes

Table 5 gives the number of optimal attribute subsets based on different clustering algorithms. Through Table 5, we have the following findings.

1) In most cases, all algorithms can remove some candidate attributes. However, the entire conditional attributes obtained by

Table 5
The number of optimal attribute subsets of different clustering algorithms.

Data set	Baseline	USQR	UEBR	UFRFS	FEUAR	LS			FSFS			SPEC			EUIAR		
		All	All	All	All	*k*-Means	*k*-Medoids	FCM	*k*-Means	*k*-Medoids	FCM	*k*-Means	*k*-Medoids	FCM	*k*-Means(λ)	*k*-Medoids(λ)	FCM(λ)
audi	69	14	14	41	23	2	2	9	45	1	49	68	66	67	64 (–)	64 (–)	63 (–)
bala	4	4	4	4	4	2	2	1	2	3	1	3	3	1	2 (–)	3 (–)	1 (–)
brea	9	9	9	9	9	3	1	2	5	1	1	2	2	2	2 (–)	2 (–)	2 (–)
monk	6	6	6	6	6	3	3	2	1	1	1	5	2	2	1 (–)	1 (–)	1 (–)
mush	22	14	14	18	15	17	21	21	9	19	19	19	19	21	11 (–)	19 (–)	17 (–)
soyb	35	6	6	9	5	13	11	8	34	21	21	33	21	21	28 (–)	30 (–)	32 (–)
spect	22	18	17	20	20	13	6	8	12	8	10	2	1	17	2 (–)	1 (–)	7 (–)
tic	9	8	8	9	8	1	1	5	3	1	3	8	5	8	2 (–)	2 (–)	3 (–)
vote	16	15	15	16	16	11	11	11	15	15	15	14	15	14	13 (–)	15 (–)	14 (–)
derm	34	11	13	19	34	19	33	32	27	31	33	32	33	27	29 (1.7)	33 (0.8)	32 (1.0)
diab	8	8	8	8	8	2	2	2	6	5	7	4	6	4	7 (0.7)	6 (0.5)	7 (0.9)
iono	33	18	18	12	33	10	25	9	5	4	4	33	33	33	5 (1.3)	9 (1.9)	5 (1.3)
iris	4	4	4	4	4	2	2	2	2	3	3	1	1	2	1 (0.1)	1 (0.1)	1 (0.1)
move	90	13	13	5	90	87	86	42	56	61	19	87	84	28	11 (0.7)	10 (1.0)	4 (0.1)
wave	21	12	12	12	21	3	17	3	12	16	7	1	15	1	1 (1.2)	14 (1.4)	11 (1.2)
wine1	13	10	10	8	13	7	9	11	10	9	9	12	6	12	9 (0.8)	9 (0.5)	12 (0.1)
wine2	11	11	11	10	11	4	10	10	1	1	10	5	5	10	5 (1.6)	3 (0.8)	8 (2.0)
yeast	8	7	7	8	8	7	7	2	7	7	5	7	7	2	7 (0.5)	7 (0.9)	7 (0.4)
abalone	8	8	8	8	8	1	1	4	1	1	2	1	1	7	1 (1.0)	1 (0.1)	2 (1.5)
autos	25	11	12	16	22	4	8	22	18	8	20	13	9	23	23 (0.5)	18 (0.4)	21 (1.3)
bands	39	4	5	16	19	1	1	16	4	4	2	7	6	6	27 (0.3)	8 (0.4)	29 (0.6)
flag	28	12	12	26	18	1	1	10	1	1	26	8	18	22	16 (0.4)	15 (1.1)	17 (0.6)
heart	13	10	10	13	13	9	8	10	11	11	11	12	7	7	12 (2.0)	12 (1.5)	7 (0.7)
hepa	19	12	13	18	19	11	11	8	12	1	12	14	13	18	13 (1.9)	9 (1.3)	15 (1.0)
horse	27	3	3	21	6	4	4	12	3	17	12	4	3	1	4 (0.1)	1 (0.1)	1 (0.1)
labor	16	9	7	14	14	1	4	8	13	8	14	12	3	12	13 (0.7)	11 (0.7)	14 (0.8)
sick	29	20	18	27	27	14	8	2	23	27	2	1	1	2	13 (2.0)	6 (1.9)	5 (1.6)
Average	22.9	10.3	10.3	14.0	17.6	9.3	10.9	10.1	12.5	10.6	11.8	15.1	14.3	13.7	11.9	11.5	12.5



(a) abalone (b) autos (c) bands

(d) flag (e) heart (f) hepa

(g) horse (h) labor (i) sick

Fig. 2. The average clustering accuracy varies with λ and the number of attributes on mixed data sets

USQR, UEBR, UFRFS, and FEUAR on some data sets such as bala, brea, monk, iris, and abalone. This shows that these four attribute reduction algorithms cannot effectively remove redundant or irrelevant attributes on some data sets.

2) In most cases, for different clustering algorithms, the same feature selection algorithm requires different attribute subsets to produce the best clustering performance. For example, for EUIAR, the number of optimal attribute subsets on data sets spect for *k*-Means, *k*-Mediods, and FCM are 2, 1, and 7, respectively.
3) For some data sets, the number of optimal attribute subsets of the same feature selection algorithm under different clustering algorithms may be the same. For example, EUIAR has the same number of optimal attribute subsets for these three clustering algorithms on data sets brea, monk, iris, and yeast.
4) In terms of the average number of attribute subsets, EUIAR obtains relatively small values in most cases. For example, for *k*-Means, the average number of attribute subsets of EUIAR is 11.9, which is smaller than URFFS, FEUAR, FSFS, and SPEC but slightly larger than other attribute reduction algorithms.

5.4.2. Accuracy

Tables 6 –8 respectively give a comparison of the optimal clustering accuracy based on *k*-Means, *k*-Mediods, and FCM. The numbers in bold face denote the best accuracy of different selection algorithms. Through Tables 6–8, the corresponding analyses are as follows.

1) From Tables 6 to 8, it can be seen that EUIAR can improve or maintain the clustering accuracy of the original data on all data sets.
2) Table 6 shows that EUIAR has 17 data sets to achieve the best clustering accuracy. However, for USQR, UEBR, UFRFS, FEUAR, LS, FSFS, and SPEC, only 0, 0, 1, 0, 7, 2, and 4 data sets achieve the best accuracy.
3) Table 7 shows that EUIAR has 22 data sets to achieve the best clustering accuracy. However, for USQR, UEBR, UFRFS, FEUAR, LS, FSFS, and SPEC, only 0, 1, 1, 1, 12, 5, and 10 data sets achieve the best accuracy.
4) Through Table 8, we can see that EUIAR achieves the best clustering accuracy on 18 data sets. However, for USQR, UEBR,

Table 6
The clustering results (Acc% ± std%) of the reduced data on *k*-Means.

Data set	Baseline	USQR	UEBR	UFRFS	FEUAR	LS	FSFS	SPEC	EUIAR
audi	42.61 ± 3.67	35.22 ± 2.28	35.22 ± 2.88	39.96 ± 2.74	35.62 ± 2.53	**48.19 ± 1.75**	44.73 ± 3.32	43.94 ± 2.64	46.15 ± 2.22
bala	51.81 ± 2.60	–	–	–	–	**54.34 ± 2.48**	51.57 ± 2.24	54.19 ± 2.97	52.83 ± 2.61
brea	63.88 ± 8.68	–	–	–	–	64.51 ± 9.19	58.11 ± 8.67	**72.38 ± 0.00**	68.46 ± 8.26
monk	55.46 ± 4.44	–	–	–	–	56.25 ± 6.91	**63.89 ± 0.00**	55.00 ± 8.39	**63.89 ± 0.00**
mush	75.89 ± 13.87	71.15 ± 8.99	74.32 ± 7.67	74.42 ± 14.80	75.61 ± 7.49	78.42 ± 7.49	74.40 ± 0.00	75.13 ± 12.81	**83.00 ± 0.00**
soyb	91.49 ± 13.86	69.36 ± 9.99	74.26 ± 6.06	62.13 ± 8.85	43.83 ± 3.91	**100.00 ± 0.00**	94.04 ± 12.60	96.17 ± 8.62	95.74 ± 4.49
spect	56.63 ± 0.79	58.43 ± 0.00	59.33 ± 0.36	56.67 ± 0.95	57.12 ± 1.09	61.42 ± 0.00	58.54 ± 4.00	**67.19 ± 4.64**	64.94 ± 0.19
tic	62.39 ± 4.14	60.30 ± 4.71	63.04 ± 4.17	–	56.25 ± 1.40	**68.60 ± 2.82**	67.05 ± 1.84	56.42 ± 1.33	64.90 ± 4.14
vote	88.05 ± 0.00	85.06 ± 0.00	85.06 ± 0.00	–	–	88.28 ± 0.00	85.06 ± 0.00	85.75 ± 0.00	**88.97 ± 0.00**
derm	70.19 ± 11.05	60.66 ± 7.27	65.85 ± 6.64	**77.16 ± 7.16**	65.52 ± 15.91	73.55 ± 8.27	74.43 ± 14.05	72.30 ± 10.25	76.97 ± 9.33
diab	66.45 ± 1.11	–	–	–	–	67.06 ± 0.00	66.32 ± 0.56	67.06 ± 0.00	**69.93 ± 1.98**
iono	71.23 ± 0.00	68.66 ± 0.00	70.37 ± 0.00	72.11 ± 3.65	71.23 ± 0.00	71.23 ± 0.00	74.93 ± 0.00	70.66 ± 0.00	**83.28 ± 4.47**
iris	82.40 ± 13.21	–	–	–	–	**96.00 ± 0.00**	80.00 ± 0.00	**96.00 ± 0.00**	**96.00 ± 0.00**
move	44.00 ± 3.00	42.94 ± 1.96	40.47 ± 1.49	30.44 ± 1.30	44.25 ± 1.72	45.11 ± 2.08	45.14 ± 1.36	45.00 ± 1.85	**46.89 ± 2.17**
wave	50.12 ± 0.00	50.92 ± 0.00	50.20 ± 0.00	50.58 ± 0.00	50.12 ± 0.00	52.72 ± 0.00	51.42 ± 0.00	51.80 ± 0.19	**52.83 ± 0.09**
wine1	95.28 ± 0.29	95.73 ± 0.71	94.94 ± 0.00	93.26 ± 0.00	95.17 ± 0.29	95.84 ± 0.29	94.94 ± 0.00	94.49 ± 0.58	**97.19 ± 0.00**
wine2	26.10 ± 1.55	–	–	28.71 ± 2.39	–	29.30 ± 2.37	29.05 ± 2.26	29.56 ± 0.90	**33.02 ± 1.75**
yeast	39.45 ± 2.24	39.43 ± 2.29	38.14 ± 1.08	–	–	39.99 ± 1.35	39.95 ± 3.31	39.49 ± 2.42	**40.65 ± 2.68**
abalone	15.18 ± 0.49	–	–	–	–	19.35 ± 0.04	**19.37 ± 0.05**	19.37 ± 0.04	**19.37 ± 0.05**
autos	41.61 ± 3.51	41.85 ± 3.68	43.90 ± 5.15	40.88 ± 3.40	39.85 ± 3.67	**44.20 ± 0.47**	43.41 ± 3.09	41.41 ± 2.21	43.61 ± 2.32
bands	56.89 ± 2.69	58.76 ± 0.00	58.59 ± 0.58	54.24 ± 1.19	57.18 ± 4.19	64.22 ± 0.00	63.99 ± 0.36	64.22 ± 0.00	**67.04 ± 2.43**
flag	28.81 ± 2.73	28.61 ± 1.31	29.90 ± 2.42	29.85 ± 2.72	27.42 ± 3.01	**42.47 ± 0.27**	35.05 ± 0.00	35.31 ± 1.50	35.98 ± 3.72
heart	77.59 ± 6.44	63.19 ± 7.49	65.19 ± 6.64	–	–	76.19 ± 6.40	76.22 ± 6.40	75.33 ± 8.47	**79.63 ± 0.00**
hepa	61.29 ± 0.00	60.45 ± 2.32	59.87 ± 0.41	60.65 ± 0.00	60.65 ± 2.04	61.94 ± 0.00	69.35 ± 5.10	61.29 ± 0.00	**72.90 ± 0.00**
horse	53.53 ± 0.00	65.95 ± 1.98	62.20 ± 3.02	56.47 ± 5.04	62.04 ± 3.17	57.61 ± 7.16	60.76 ± 6.45	66.33 ± 0.09	**69.02 ± 3.94**
labor	77.02 ± 9.73	70.35 ± 7.37	54.21 ± 1.54	78.07 ± 11.67	79.30 ± 15.42	84.21 ± 0.00	81.75 ± 11.73	79.65 ± 6.57	**88.77 ± 5.44**
sick	73.98 ± 9.52	76.25 ± 6.44	74.82 ± 6.46	73.83 ± 7.12	77.59 ± 9.13	84.42 ± 7.86	83.48 ± 8.91	**93.90 ± 0.00**	92.61 ± 0.52
Average	–	–	–	–	–	63.90 ± 2.49	62.48 ± 3.57	63.31 ± 2.83	**66.47 ± 2.33**

Table 7
The clustering results (Acc% ± std%) of the reduced data on k-Medoids.

Data set	Baseline	USQR	UEBR	UFRFS	FEUAR	LS	FSFS	SPEC	EUIAR
audi	39.29 ± 0.28	33.89 ± 0.37	31.42 ± 1.55	34.51 ± 0.42	33.05 ± 0.21	**50.44 ± 0.00**	46.46 ± 0.00	39.47 ± 0.19	41.95 ± 0.50
bala	51.49 ± 6.06	–	–	–	–	**54.67 ± 2.93**	53.95 ± 4.07	54.26 ± 4.23	52.85 ± 7.04
brea	51.40 ± 0.00	–	–	–	–	51.40 ± 0.00	53.08 ± 0.36	**72.38 ± 0.00**	**72.38 ± 0.00**
monk	56.94 ± 0.00	–	–	–	–	56.94 ± 0.00	**70.56 ± 9.73**	50.56 ± 0.72	**70.56 ± 9.73**
mush	78.60 ± 6.53	71.64 ± 10.05	70.90 ± 11.12	77.73 ± 5.74	71.75 ± 10.57	81.28 ± 6.94	82.24 ± 6.15	78.49 ± 6.46	**80.17 ± 6.45**
soyb	100.00 ± 0.00	72.55 ± 3.08	72.55 ± 3.08	58.72 ± 3.64	46.81 ± 1.00	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**
spect	52.06 ± 0.00	59.18 ± 0.00	52.43 ± 0.00	50.56 ± 0.00	50.56 ± 0.00	59.40 ± 5.61	61.57 ± 5.61	**66.29 ± 0.00**	64.79 ± 0.00
tic	57.89 ± 8.40	65.99 ± 5.22	51.13 ± 0.26	–	53.03 ± 0.00	**67.27 ± 3.45**	65.93 ± 3.45	54.94 ± 1.29	**67.93 ± 3.23**
vote	86.16 ± 0.24	83.91 ± 0.24	83.86 ± 0.24	–	–	**88.28 ± 0.00**	84.05 ± 0.19	83.91 ± 0.24	**87.22 ± 0.78**
derm	92.62 ± 0.00	64.48 ± 0.00	69.56 ± 0.35	**71.31 ± 0.00**	**92.62 ± 0.00**	**92.62 ± 0.00**	72.95 ± 0.00	**92.62 ± 0.00**	**92.62 ± 0.00**
diab	66.33 ± 0.66	–	–	–	–	67.06 ± 0.00	66.69 ± 1.15	66.93 ± 0.00	**71.54 ± 3.18**
iono	70.94 ± 0.00	68.09 ± 0.00	70.37 ± 0.00	70.83 ± 3.53	70.94 ± 0.00	71.23 ± 0.00	74.93 ± 0.00	70.37 ± 0.00	**75.67 ± 4.85**
iris	90.27 ± 0.34	–	–	–	–	**96.00 ± 0.00**	84.00 ± 0.00	**96.00 ± 0.00**	**96.00 ± 0.00**
move	45.78 ± 0.88	44.42 ± 1.11	40.94 ± 0.77	31.14 ± 0.72	45.69 ± 1.17	47.19 ± 1.34	47.67 ± 1.52	**47.97 ± 1.24**	**47.92 ± 1.73**
wave	52.83 ± 0.28	58.72 ± 3.34	57.07 ± 2.80	52.15 ± 2.42	53.02 ± 0.51	58.60 ± 6.05	57.63 ± 7.12	58.36 ± 6.06	**61.30 ± 5.54**
wine1	92.70 ± 0.00	92.36 ± 0.29	**95.51 ± 0.00**	93.26 ± 0.00	92.70 ± 0.00	**95.51 ± 0.00**	94.38 ± 0.00	93.60 ± 2.84	**94.94 ± 0.00**
wine2	27.78 ± 1.65	–	–	27.68 ± 2.00	–	28.26 ± 1.89	29.08 ± 2.15	30.30 ± 2.01	**31.71 ± 1.19**
yeast	39.25 ± 2.37	36.76 ± 1.77	36.06 ± 1.84	–	–	38.72 ± 2.13	39.20 ± 2.26	39.06 ± 1.64	**40.53 ± 1.19**
abalone	15.72 ± 0.65	–	–	–	–	**19.30 ± 0.00**	**19.30 ± 0.00**	**19.30 ± 0.00**	**19.30 ± 0.00**
autos	43.41 ± 0.00	42.20 ± 3.46	40.98 ± 0.00	42.93 ± 0.00	44.88 ± 0.00	**44.49 ± 0.31**	43.66 ± 2.76	43.90 ± 0.00	**45.37 ± 0.00**
bands	56.50 ± 0.00	57.06 ± 0.00	56.91 ± 2.24	54.80 ± 0.00	56.31 ± 0.00	**64.22 ± 0.00**	62.69 ± 2.46	**64.22 ± 0.00**	**64.22 ± 0.00**
flag	30.77 ± 0.97	29.07 ± 0.65	29.02 ± 0.77	32.53 ± 1.47	26.60 ± 2.99	**42.78 ± 0.00**	35.05 ± 0.00	35.46 ± 1.30	35.77 ± 2.99
heart	72.74 ± 5.14	58.52 ± 0.00	58.52 ± 0.00	–	–	78.52 ± 0.00	77.78 ± 0.00	**78.89 ± 0.00**	**78.89 ± 0.00**
hepa	61.94 ± 0.00	61.94 ± 0.00	61.94 ± 0.00	60.52 ± 2.99	61.94 ± 0.00	65.16 ± 3.40	**66.45 ± 0.00**	63.23 ± 2.72	**65.81 ± 0.00**
horse	53.80 ± 0.00	62.83 ± 3.23	60.95 ± 1.98	53.80 ± 0.00	65.33 ± 2.64	55.16 ± 0.00	55.43 ± 0.00	66.30 ± 0.00	**66.58 ± 0.00**
labor	74.56 ± 2.77	72.63 ± 5.44	56.14 ± 0.00	75.44 ± 0.00	74.56 ± 2.77	84.74 ± 1.66	**87.72 ± 0.00**	78.95 ± 0.00	**87.72 ± 0.00**
sick	80.43 ± 9.41	75.29 ± 6.77	75.08 ± 8.73	82.90 ± 7.74	84.29 ± 7.13	87.62 ± 0.00	87.62 ± 0.00	**93.90 ± 0.00**	92.69 ± 0.53
Average	60.82 ± 1.73	–	–	–	–	64.70 ± 1.32	63.71 ± 1.81	64.43 ± 1.15	**66.90 ± 1.81**

Table 8
The clustering results (Acc% ± std%) of the reduced data on FCM.

Data set	Baseline	USQR	UEBR	UFRFS	FEUAR	LS	FSFS	SPEC	EUIAR
audi	34.96 ± 0.29	40.27 ± 0.00	39.91 ± 0.50	34.91 ± 0.33	34.96 ± 0.00	46.77 ± 3.55	47.88 ± 1.28	35.35 ± 2.98	**47.92 ± 2.69**
bala	52.93 ± 8.25	–	–	–	–	**54.88 ± 0.00**	**54.88 ± 0.00**	**54.88 ± 0.00**	**54.88 ± 0.00**
brea	51.40 ± 0.00	–	–	–	–	51.40 ± 0.00	52.80 ± 0.00	**72.38 ± 0.00**	**72.38 ± 0.00**
monk	50.00 ± 0.00	–	–	–	–	50.00 ± 0.00	**63.89 ± 0.00**	50.00 ± 0.00	**63.89 ± 0.00**
mush	83.49 ± 0.00	79.97 ± 0.00	79.97 ± 0.00	83.10 ± 0.00	79.90 ± 0.00	83.42 ± 0.00	83.47 ± 0.00	83.49 ± 0.00	**84.21 ± 0.00**
soyb	100.00 ± 0.00	70.21 ± 0.00	70.21 ± 0.00	50.21 ± 1.49	44.68 ± 0.00	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**
spect	60.52 ± 0.71	61.05 ± 0.00	61.42 ± 0.00	60.22 ± 0.95	59.93 ± 0.00	65.17 ± 0.00	64.04 ± 0.00	62.17 ± 0.00	**69.66 ± 0.00**
tic	64.10 ± 3.93	62.12 ± 6.24	59.62 ± 6.48	–	55.33 ± 2.25	**67.43 ± 0.00**	65.03 ± 0.00	54.76 ± 2.08	65.97 ± 0.00
vote	88.05 ± 0.00	85.52 ± 0.00	85.52 ± 0.00	–	–	88.28 ± 0.00	85.52 ± 0.00	85.98 ± 0.00	88.28 ± 0.00
derm	70.87 ± 6.16	50.41 ± 0.23	67.08 ± 0.75	60.19 ± 3.64	69.73 ± 6.07	77.16 ± 0.14	74.56 ± 5.04	77.60 ± 0.00	77.84 ± 0.20
diab	66.67 ± 0.00	–	–	–	–	66.67 ± 0.00	66.21 ± 2.68	66.41 ± 0.00	**71.68 ± 0.11**
iono	70.94 ± 0.00	68.09 ± 0.00	70.37 ± 0.00	67.52 ± 0.00	70.94 ± 0.00	71.23 ± 0.00	74.93 ± 0.00	70.94 ± 0.00	**84.05 ± 0.00**
iris	89.33 ± 0.00	–	–	–	–	**96.67 ± 0.00**	83.40 ± 0.21	**96.67 ± 0.00**	96.00 ± 0.00
move	15.61 ± 1.21	17.31 ± 2.53	19.42 ± 0.75	31.58 ± 1.08	16.61 ± 1.54	39.56 ± 1.34	**42.36 ± 0.67**	40.86 ± 0.65	38.44 ± 0.81
wave	49.30 ± 0.00	50.81 ± 1.56	51.98 ± 0.00	50.08 ± 0.01	49.30 ± 0.00	52.85 ± 0.01	57.77 ± 3.46	51.44 ± 0.00	**62.01 ± 5.95**
wine1	94.94 ± 0.00	**96.63 ± 0.00**	94.94 ± 0.00	92.70 ± 0.00	94.94 ± 0.00	94.94 ± 0.00	93.82 ± 0.00	94.38 ± 0.00	**96.63 ± 0.00**
wine2	30.06 ± 1.44	–	–	30.83 ± 1.07	–	30.61 ± 0.55	35.02 ± 2.37	28.88 ± 1.48	**36.14 ± 0.33**
yeast	32.79 ± 0.54	32.61 ± 0.29	32.30 ± 0.42	–	–	31.87 ± 0.00	32.22 ± 0.23	31.87 ± 0.00	**33.15 ± 0.29**
abalone	15.48 ± 0.88	–	–	–	–	15.50 ± 0.71	15.39 ± 0.84	15.08 ± 0.61	15.72 ± 0.83
autos	46.20 ± 0.95	39.02 ± 1.26	39.66 ± 1.75	45.02 ± 0.24	45.56 ± 0.57	48.44 ± 0.80	**48.73 ± 0.15**	46.49 ± 0.76	48.20 ± 0.31
bands	55.50 ± 0.30	56.69 ± 0.00	57.06 ± 0.00	54.61 ± 0.00	55.97 ± 1.67	56.69 ± 0.00	62.34 ± 0.00	**64.22 ± 0.00**	63.90 ± 0.20
flag	36.03 ± 2.01	31.34 ± 0.33	31.60 ± 0.81	35.57 ± 1.50	34.12 ± 2.92	36.24 ± 0.25	34.95 ± 0.93	38.51 ± 0.73	**42.37 ± 0.22**
heart	79.26 ± 0.00	69.67 ± 0.12	69.63 ± 0.00	–	–	79.26 ± 0.00	78.89 ± 0.00	80.00 ± 0.00	**82.30 ± 0.16**
hepa	61.94 ± 0.00	59.35 ± 0.00	59.35 ± 0.00	61.94 ± 0.00	61.94 ± 0.00	61.94 ± 0.00	**69.03 ± 0.00**	61.94 ± 0.00	**69.03 ± 0.00**
horse	51.96 ± 0.11	60.33 ± 0.00	60.33 ± 0.00	52.99 ± 0.00	60.33 ± 0.00	53.53 ± 0.00	55.43 ± 0.00	66.30 ± 0.00	**66.58 ± 0.00**
labor	74.91 ± 15.16	63.86 ± 1.69	52.98 ± 1.38	82.81 ± 5.35	83.68 ± 4.97	87.19 ± 5.61	85.96 ± 2.98	86.49 ± 3.79	**88.42 ± 3.22**
sick	63.97 ± 0.00	68.32 ± 0.00	68.32 ± 0.00	63.97 ± 0.00	63.97 ± 0.00	84.04 ± 0.00	68.32 ± 0.00	93.45 ± 0.00	91.30 ± 0.00
Average	58.93 ± 1.55	–	–	–	–	62.66 ± 0.48	62.85 ± 0.77	63.35 ± 0.48	**67.07 ± 0.57**

UFRFS, FEUAR, LS, FSFS and SPEC, only 1, 0, 0, 0, 4, 6, and 5 data sets achieve the best clustering accuracy.
5) From the perspective of average clustering accuracy, EUIAR achieves the maximal value in all three clustering algorithms.

Both the replacement and discretization of data values may bring changes to the data structure, which in turn leads to the loss of information. On data sets containing numerical attributes, USQR and UEBR show relatively poor clustering accuracy. This is because the numerical attributes were discretized before the experiment. In addition, for data sets containing nominal attributes, replacing the nominal attributes with different integer values will also affect the clustering performance of LS, FSFS, and SPEC. However, for EUIAR, neither replacement nor discretization needs to be done, so that more real information about the data can be retained. Therefore, EUIAR has relatively better clustering results. In addition, although FEUAR and URFFS does not require corresponding preprocessing, it does not get better clustering results. This may be because it only considers the relevance of features, and does not consider the redundancy and interactivity between features.

EUIAR can get a relatively small subset of attributes and improve or maintain clustering accuracy. Therefore, EUIAR is suitable for feature selection on mixed attribute data and the performance of clustering algorithms can be improved on reduced data.

5.5. Parameter sensitivity

The threshold λ plays an important role in EUIAR. It can be used as a parameter to control the fuzzy granularity of data analysis. Different subsets of attribute can be obtained at different levels of granularity. Then, different subset of attributes is can be used for different clustering tasks.

The average accuracy of the three clustering algorithms varies with the parameters λ and the number of selected attributes on numerical and mixed data sets as shown in Figs. 1 and 2. Obviously, it can be seen that the clustering performance is different on different data sets. Through Figs. 1 and 2, the specific analysis is as follows.

1) As far as the number of selected attributes is concerned, in most data sets, as the number of attributes increases, the clustering accuracy increases rapidly at first, and then remains unchanged or even decreases, such as data sets iono, move, bands, flag, horse, and sick. This is because more and more redundant or irrelevant attributes are selected, which cannot provide new information to the clustering task or even mislead the clustering task.
2) For data sets iono, iris, move, abalone, bands, hepa, horse, and sick, the clustering algorithm can maintain or improve the original clustering accuracy on reduced data with a small number of attributes. These results show that EUIAR can select important attributes that are representative and informative.
3) It can be seen that with the increase of λ on some data sets, the average clustering accuracy does not change much, such as data sets derm, move, wine1, and yeast. However, the average clustering accuracy fluctuates greatly as λ increases, such as diab, wave, wine2, etc. It shows that λ is too small or too big to make the result optimal.
4) From Figs. 1 and 2, it can be seen that for most data sets, the optimal value can be obtained under multiple values of λ. For each data set, the appropriate value of λ can be selected to achieve better clustering performance.

Through the above analyses, it can be seen that the performance is sensitive to λ and the number of selected attributes. However, under the conditions of the appropriate λ value and the number of selected attributes, EUIAR can achieve better results in most cases. Therefore, EUIAR is feasible for unsupervised mixed attribute reduction.

6. Conclusion

This paper proposes an interactive attribute reduction method for unlabeled mixed data. This method uses mixed fuzzy similarity measures to calculate the fuzzy similarity relation matrix. This makes the proposed method suitable for mixed attribute data. In addition, it not only considers the correlation and redundancy between attributes, but also considers the interaction between attributes. Further, the corresponding algorithm EUAIR is designed, which is compared with the existing algorithms on some public data sets. The results show that the algorithm can select fewer features to maintain or improve the ability of clustering tasks. Further, the parameter sensitivity analysis results verify that EUIAR can obtain better clustering results in most cases under the appropriate λ value and the number of selected attributes. In addition, the proposed algorithm is sensitive to parameters λ. Therefore, in future work, how to determine the optimal value of the parameter is worthy of further exploration. What's more, we can further consider the idea of concept lattices [39] and bireducts [40] for unsupervised attribute reduction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank both the editors and reviewers for their valuable suggestions, which substantially improve this paper. This work was supported by the National Natural Science Foundation of China (61976182 and 62076171), the Key Techniques of Integrated Operation and Maintenance for Urban Rail Train Dispatching Control System based on Artificial Intelligence (2019YFH0097), and Sichuan Key R&D project (2020YFG0035).

References

[1] K.Y. Liu, X.B. Yang, H. Fujita, D. Liu, X. Yang, Y.H. Qian, An efficient selector for multi-granularity attribute reduction, Inf. Sci. 505 (2019) 457–472.
[2] D. Kumar, P. Maji, Rough-Bayesian approach to select class-pair specific descriptors for HEp-2 cell staining pattern recognition, Pattern Recognit. 117 (2021) 107982.
[3] J.Y. Liang, X.W. Zhao, D.Y. Li, F.Y. Cao, C.Y. Dang, Determining the number of clusters using information entropy for mixed data, Pattern Recognit. 45 (6) (2012) 2251–2265.
[4] I. Düntsch, G. Gediga, Uncertainty measures of rough set prediction, Artif. Intell. 106 (1) (1998) 109–137.
[5] J.Y. Liang, K.-S. Chin, C.Y. Dang, R.C. Yam, A new method for measuring uncertainty and fuzziness in rough set theory, Int. J. Gen. Syst. 31 (4) (2002) 331–342.
[6] J.Y. Liang, Z.Z. Shi, The information entropy, rough entropy and knowledge granulation in rough set theory, Int. J. Uncertain., Fuzziness Knowledge-Based Syst. 12 (01) (2004) 37–46.
[7] Y.H. Qian, J.Y. Liang, F. Wang, A new method for measuring the uncertainty in incomplete information systems, Int. J. Uncertain., Fuzziness Knowledge-Based Syst. 17 (06) (2009) 855–880.
[8] J. Wang, J. Liang, Y. Qian, C. Dang, Uncertainty measure of rough sets based on a knowledge granulation for incomplete information systems, Int. J. Uncertain., Fuzziness Knowledge-Based Syst. 16 (02) (2008) 233–244.
[9] Y.H. Qian, J.Y. Liang, Combination entropy and combination granulation in rough set theory, Int. J. Uncertain., Fuzziness Knowledge-Based Syst. 16 (02) (2008) 179–193.
[10] L. Sun, J. Xu, Y. Tian, Feature selection using rough entropy-based uncertainty measures in incomplete decision systems, Knowledge-Based Syst. 36 (2012) 206–216.
[11] F. Jiang, Y.F. Sui, L. Zhou, A relative decision entropy-based feature selection approach, Pattern Recognit. 48 (7) (2015) 2151–2163.

[12] J. Zhao, J.M. Liang, Z.N. Dong, D.Y. Tang, Z. Liu, Accelerating information entropy-based feature selection using rough set theory with classified nested equivalence classes, Pattern Recognit. 107 (2020) 107517.
[13] X.Y. Zhang, Y.Y. Yao, Tri-level attribute reduction in rough set theory, Expert Syst. Appl. 190 (2022) 116187.
[14] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[15] R.R. Yager, Entropy measures under similarity relations, Int. J. Gen. Syst. 20 (4) (1992) 341–358.
[16] J.S. Mi, Y. Leung §, W.Z. Wu, An uncertainty measure in partition-based fuzzy rough sets, Int. J. Gen. Syst. 34 (1) (2005) 77–90.
[17] Q.H. Hu, D.R. Yu, Z.X. Xie, J.F. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.
[18] Y.H. Qian, J.Y. Liang, W. Wei Zhi, C.Y. Dang, Information granularity in fuzzy binary GrC model, IEEE Trans. Fuzzy Syst. 19 (2) (2010) 253–264.
[19] X. Zhang, C.L. Mei, D.G. Chen, J.H. Li, Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy, Pattern Recognit. 56 (2016) 1–15.
[20] C.Z. Wang, Y. Huang, M.W. Shao, D.G. Chen, Uncertainty measures for general fuzzy relations, Fuzzy Sets Syst. 360 (2019) 82–96.
[21] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognit. Lett. 27 (5) (2006) 414–423.
[22] S. An, Q.H. Hu, D.R. Yu, Fuzzy entropy based max-relevancy and min-redundancy feature selection, in: 2008 IEEE International Conference on Granular Computing, IEEE, 2008, pp. 101–106.
[23] J.H. Dai, Q. Xu, Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification, Appl. Soft Comput. 13 (1) (2013) 211–221.
[24] Y.J. Lin, Q.H. Hu, J.H. Liu, J.J. Li, X.D. Wu, Streaming feature selection for multilabel learning based on fuzzy mutual information, IEEE Trans. Fuzzy Syst. 25 (6) (2017) 1491–1507.
[25] J.H. Dai, J.L. Chen, Feature selection via normative fuzzy information weight with application in biological data classification, Appl. Soft Comput. 92 (2020) 106–299.
[26] C.Z. Wang, Y. Huang, W.P. Ding, Z.H. Cao, Attribute reduction with fuzzy rough self-information measures, Inf. Sci. 49 (5) (2021) 68–86.
[27] L. Sun, L.Y.Y. Wang, W.P. Ding, Y.H. Qian, J.C. Xu, Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. 29 (1) (2020) 19–33.
[28] A. Ganivada, S.S. Ray, S.K. Pal, Fuzzy rough sets, and a granular neural network for unsupervised feature selection, Neural Netw. 48 (2013) 91–108.
[29] N. Mac Parthaláln, R. Jensen, Unsupervised fuzzy-rough set-based dimensionality reduction, Inf. Sci. 229 (2013) 106–121.
[30] Z.L. Zeng, H.J. Zhang, R. Zhang, C.X. Yin, A novel feature selection method considering feature interaction, Pattern Recognit. 48 (8) (2015) 2656–2666.
[31] Z. Yuan, H.M. Chen, P. Xie, P.F. Zhang, J. Liu, T.R. Li, Attribute reduction methods in fuzzy rough set theory: an overview, comparative experiments, and new directions, Appl. Soft Comput. (2021) 107353.
[32] Z. Yuan, H.M. Chen, T.R. Li, J. Liu, S. Wang, Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection, Fuzzy Sets Syst. 421 (2021) 1–28.
[33] C. Velayutham, K. Thangavel, Unsupervised quick reduct algorithm using rough set theory, J. Electron. Sci. Technol. 9 (3) (2011) 193–201.
[34] C. Velayutham, K. Thangavel, A novel entropy based unsupervised feature selection algorithm using rough set theory, in: IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012), IEEE, 2012, pp. 156–161.
[35] X.F. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Advances in Neural Information Processing Systems, 2006, pp. 507–514.
[36] P. Mitra, C. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 301–312.
[37] Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 1151–1157.
[38] P.F. Zhu, W.C. Zhu, Q.H. Hu, C.Q. Zhang, W.M. Zuo, Subspace clustering guided unsupervised feature selection, Pattern Recognit. 66 (2017) 364–374.
[39] J. Medina, Relating attribute reduction in formal, object-oriented and property-oriented concept lattices, Comput. Math. Appl. 64 (6) (2012) 1992–2002.
[40] M.J. Benítez-Caballero, J. Medina, E. Ramírez-Poussa, D. Ślęzak, Bireducts with tolerance relations, Inf. Sci. 435 (2018) 26–39.

Zhong Yuan received the B.Sc. degree in mathematics from Sichuan Minzu College, Kangding, China, in 2015. He received the M.Sc. in mathematics from Sichuan Normal University, Chengdu, China, in 2018. He is currently pursuing the Ph.D. degree with Southwest Jiaotong University, Chengdu, China. His research interests include rough sets, granular computing, and data mining.

Hongmei Chen received the M.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2000. She received the Ph.D. degree from the Southwest Jiaotong University, Chengdu, China, in 2013. She is currently a professor at the School of Computing and Artificial Intelligence, Southwest Jiaotong University. Her research interests include the areas of data mining, pattern recognition, fuzzy sets, and rough sets.

Tianrui Li received the BS, MS, and Ph.D. degrees from the Southwest Jiaotong University, Chengdu, China, in 1992, 1995, and 2002, respectively. He was a postdoctoral researcher at the Belgian Nuclear Research Centre (SCK.CEN), Belgium from 2005 to 2006, a visiting professor at Hasselt University, Belgium, in 2008 and the University of Technology, Sydney, Australia in 2009. Currently, he is a professor and the director of the Key Lab of Cloud Computing and Intelligent Technology, Southwest Jiaotong University, China. His research interests include data mining and knowledge discovery, granular computing and rough sets, cloud computing. Since 2000, he has coedited two textbooks, six proceedings, three special issues of international journal and published more than 90 research papers in refereed journals and conferences. He is the vice chair of IEEE CIS Chengdu Chapter and the area editor of the International Journal of Computational Intelligence Systems (IJCIS). He has served as ISKE2007, ISKE2008, ISKE2009, ISKE2010, ISKE2011 program chairs, IEEE GrC 2009 program vice chair and RSKT2008, FLINS2010 organizing chairs and has been a reviewer for several leading academic journals. He is a senior member of the IEEE.