Information Sciences 572 (2021) 67–87

Contents lists available at ScienceDirect

Information Sciences

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

# Unsupervised attribute reduction for mixed data based on fuzzy rough sets

Zhong Yuan [a,b], Hongmei Chen [a,b,*], Tianrui Li [a,b], Zeng Yu [a,b], Binbin Sang [a,b], Chuan Luo [c]

[a] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
[b] National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China
[c] College of Computer Science, Sichuan University, Chengdu 610065, China

ARTICLE INFO

*Article history:*
Received 27 June 2020
Received in revised form 19 April 2021
Accepted 27 April 2021
Available online 2 May 2021

*Keywords:*
Fuzzy rough sets
Fuzzy dependency
Unsupervised attribute reduction
Mixed data

ABSTRACT

Unsupervised attribute reduction becomes very challenging due to a lack of decision information, which is to select a subset of attributes that can maintain learning ability without decision information. However, most of the existing unsupervised attribute reduction methods are proposed for numerical or nominal attributes, and little research has been done on unsupervised mixed attribute reduction methods. In view of this, this paper proposes a generalized unsupervised mixed attribute reduction model based on fuzzy rough sets. First, based on all single attribute subsets, the significance is defined to indicate the importance of a candidate attribute. Then, a specific fuzzy rough-based unsupervised attribute reduction (FRUAR) algorithm is designed. Finally, the proposed algorithm is compared with the existing algorithms by using thirty public data sets. Experimental results show that the algorithm FRUAR can select fewer attributes to maintain or improve the performance of learning algorithms, and it is suitable for mixed attribute data.



## 1. Introduction

Attribute reduction (also called feature selection) has become an effective method for processing high-dimensional data. Its goal is to find a minimum attribute subset in the high-dimensional attribute space to improve or maintain the learning performance of the original data [49]. According to whether there is decision (or label) information, the existing attribute reduction methods can be divided into (1) supervised, (2) semi-supervised, and (3) unsupervised method [37]. Among them, for the supervised and semi-supervised methods, all or part of the decision information of objects must be known in advance. However, obtaining the decision information of an object is actually a challenging task. Therefore, how to extract the most discriminative attributes from high-dimensional data without decision information is meaningful in real-life applications, which attracts more and more researchers to study unsupervised attribute reduction [49,37].

Unsupervised attribute reduction methods can be generally divided into (1) filter-based [8], (2) wrapper-based [14], and (3) embedded-based method [38]. The filter-based method uses one or several different evaluation indicators to evaluate the importance of a single attribute or a subset of attributes, and then sorts them according to their importance for selecting attributes with higher importance [8]. The wrapper-based method involves learning algorithm in attribute reduction meth-

* Corresponding author at: School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China.
*E-mail addresses:* yuanzhong2799@foxmail.com (Z. Yuan), hmchen@swjtu.edu.cn (H. Chen), trli@swjtu.edu.cn (T. Li), zyu@swjtu.edu.cn (Z. Yu), sangbinbin@my.swjtu.edu.cn (B. Sang), cluo@scu.edu.cn (C. Luo).

https://doi.org/10.1016/j.ins.2021.04.083
0020-0255/

ods and uses the classification error rate or performance accuracy as the evaluation criterion of attributes [50]. The embedded-based method is a built-in attribute reduction mechanism. It embeds attribute reduction into a learning algorithm and uses the learning algorithm to guide attribute evaluation [38].

Among the above three unsupervised methods, the filter-based method usually reduces attributes based on four evaluation methods: distance-based, information-based, correlation coefficient-based, and consistency-based [20]. Among them, information-based and consistency-based methods are generally applied to nominal attributes, and distance-based and correlation coefficient-based methods are used for numerical attributes [20]. However, in real-life applications, data may contain mixed attribute. Therefore, some of the above methods are essentially applicable to data sets with single attribute type (i.e., nominal or numerical).

Recently, the classical rough set theory has been successfully applied to attribute reduction [11,45,47], group decision making [31], uncertainty reasoning [6], etc. However, a learning model is constructed based on Boolean equivalence relations in the classical rough set, which is only applicable to nominal attribute data. In fact, nominal, numerical, and fuzzy attributes often coexist in real-life databases. Discretization is one of the feasible methods to deal with numerical attributes, which convert the numerical attributes into nominal attributes. However, discretization is often an important cause of information loss when dealing with complex data [46].

In response to the above problems, Dubois and Prade first proposed a fuzzy rough set model [12,13]. It provides an effective tool to overcome the discretization problem and can be directly applied to numerical or mixed attribute data. In the fuzzy rough set model, the fuzzy relation is defined to measure the similarity between objects, and the numerical attribute values no longer need to be discretized. The fuzzy rough set has been the focus of much attention in recent years. Inspired by the original fuzzy rough set model, a series of extensions of the fuzzy rough set were proposed [40,25]. Fuzzy rough set theory has been successfully applied to many fields, such as attribute reduction [18,7,30,27,34], rule extraction [39], and classification tree [2].

Generally, the attribute reduction methods using fuzzy rough sets can be divided into (1) fuzzy dependency function-based [21], (2) fuzzy discernibility matrix-based [5], and (3) fuzzy uncertainty measure-based [20]. Among them, dependency function-based and fuzzy uncertainty measure-based methods mainly consider the correlation of attribute subset or single attribute to reduce attributes. In some methods, the correlation between attribute subsets and decisions is used to define the significance of attributes. For example, Jensen and Shen first extended the dependency function in the classic rough set to the fuzzy rough set and proposed a related feature selection algorithm [21]. Hu et al. introduced a novel fuzzy rough set to design the heuristic algorithm for ensuring the convergence of the algorithm [18]. Wang et al. introduced distance metric to fuzzy rough sets, and then proposed a new method of attribute reduction based on fuzzy dependency [35]. In [36], Wang et al. introduced a type of fuzzy similarity relation to control the similarity between objects, and then proposed a novel attribute reduction based on the fuzzy dependency for categorical data. Hu et al. proposed fuzzy information entropy to characterize uncertainty in fuzzy approximation space and applied it to supervised feature selection [20]. In these algorithms, fuzzy dependency functions or fuzzy information measures are employed to indicate the correlation between attribute subsets and decisions. Further, the significance of attributes is defined by the difference in the correlation of attribute subsets. When the significance of a candidate attribute is the largest one, it is considered to be the best attribute. Furthermore, it is selected into the reduction subset. Therefore, we can find a better attribute subset through the significance of attributes. Also, the maximum Relevance and Minimum Redundancy (mRMR) method proposed in [28] is a widely studied feature selection method, which mainly considers the relevance between a candidate feature and decisions and the redundancy between the candidate feature and the selected feature subset. For example, Lin et al. introduced the mRMR method to evaluate the quality of features in multi-label learning and designed an effective multi-label feature selection algorithm [23]. The algorithms mentioned above usually output an attribute sequence sorted by attribute importance in descending order.

Although the above methods have been proven the effectiveness of fuzzy rough sets for attribute reduction, it is worth noting that all of these attribute reduction methods are supervised. As mentioned before, for the supervised method, the decision information of objects must be known in advance. There are many data in life without decision information. In reality, it is very expensive to obtain the decision information of objects. To our knowledge, there is only a small amount of work for unsupervised attribute reduction using fuzzy rough sets [16,24]. For example, Ganivada et al. used the fuzzy rough set model to study an unsupervised feature selection based on a granular neural network [16]. In [24], Mac ParthaláIn et al. introduced several unsupervised dimensionality reduction methods based on fuzzy rough sets.

Based on the above discussions, this paper proposes a generalized unsupervised mixed attribute reduction model based on fuzzy rough sets. First, the average dependency on all single attribute subsets is integrated to define the relevance of a certain subset of attributes, thereby the significance is defined to indicate the importance of a candidate attribute. Then, a specific Fuzzy Rough set-based Unsupervised Attribute Reduction (FRUAR) algorithm is designed to perform attribute reduction. Finally, the comparison and analysis of the algorithm FRUAR with existing algorithms are made using thirty public data sets. The experimental results indicate that our algorithm can select fewer attributes to maintain or improve the learning ability of classification, clustering, and outlier detection for mixed attribute data. Furthermore, the hypothesis statistical test shows that the proposed algorithm is statistically significantly different from most other existing algorithms.

The rest of this paper is organized as follows. In the second section, we introduce preliminary knowledge about fuzzy rough sets. And in the next section, we give a generalized unsupervised mixed attribute reduction model, analyze a specific unsupervised attribute reduction model, and propose a corresponding unsupervised attribute reduction algorithm. The experimental results are shown in Section 4. Finally, Section 5 gives a summary of the paper.

## 2. Preliminaries

This section reviews some definitions and symbols used in the subsequent sections of this paper on fuzzy rough sets [12,40,19].

A Fuzzy Information System (FIS) is a quadruple $FIS = (U, A, V, f)$, where $U$ is a non-empty finite set of objects, called universe of discourse (universe); $A$ is a non-empty finite set of attributes; $V$ is an union of attribute domain, i.e. $V = \bigcup_{a \in A} V_a$, where $V_a$ is the attribute domain of the attribute $a$;$f : U \times A \to V$ is an information function that satisfies for $\forall a \in A$ and $x \in U$, we have $f(x, a) \in V_a$. When $A = C \cup D$ and $C \cap D = \varnothing$, the FIS is called a Fuzzy Decision System (FDS), where $C$ denotes the condition attribute and $D$ denotes the decision attributes.

Let $U = \{x_1, x_2, \ldots, x_n\}$, if $\mathscr{A}$ is a map of $U$ to $[0, 1]$ , which is $\mathscr{A} : U \to [0, 1]$, then $\mathscr{A}$ is called a fuzzy set on $U$. For $\forall x_i \in U, \mathscr{A}(x_i)$ is called the membership function of $\mathscr{A}$, or the membership of $x_i$ for $\mathscr{A}$. The entire fuzzy sets on $U$ is recorded as $\mathscr{F}(U)$. Thus, $P(U) \subseteq \mathscr{F}(U)$, where $P(U)$ is the power set of $U$. The fuzzy set is denoted as $\mathscr{A} = (\mathscr{A}(x_1), \mathscr{A}(x_2), \ldots, \mathscr{A}(x_n))$ or $\mathscr{A} = \sum_{i=1}^{n} \mathscr{A}(x_i)/x_i$.

Let $\mathscr{A}, \mathscr{B} \in \mathscr{F}(U)$. For $\forall x \in U$, some operations of fuzzy set are defined as follows.

(1) $\mathscr{A} = \mathscr{B} \Longleftrightarrow \mathscr{A}(x) = \mathscr{B}(x)$;
(2) $\mathscr{A} \subseteq \mathscr{B} \Longleftrightarrow \mathscr{A}(x) \leqslant \mathscr{B}(x)$;
(3) $(\mathscr{A} \cup \mathscr{B})(x) = \max\{\mathscr{A}(x), \mathscr{B}(x)\} = \mathscr{A}(x) \vee \mathscr{B}(x)$;
(4) $(\mathscr{A} \cap \mathscr{B})(x) = \min\{\mathscr{A}(x), \mathscr{B}(x)\} = \mathscr{A}(x) \wedge \mathscr{B}(x)$;
(5) $\mathscr{A}^c(x) = 1 - \mathscr{A}(x)$.

Let $U = \{u_1, u_2, \cdots, u_m\}$ and $V = \{v_1, v_2, \cdots, v_n\}$, the fuzzy relation $\mathscr{R}$ from $U$ to $V$ is defined as a fuzzy set $\mathscr{R} : U \times V \to [0, 1]$. For $\forall (u, v) \in U \times V$, the membership degree $\mathscr{R}(u, v)$ indicates the degree to which $u$ has a relation $\mathscr{R}$ with $v$.

In particular, the fuzzy relation from $U$ to $U$ is called the fuzzy relation on $U$.

A fuzzy relation $\mathscr{R}$ from $U$ to $V$ can be represented by a fuzzy relation matrix, i.e. $M_{\mathscr{R}} = (r_{ij})_{m \times n}$, where $r_{ij} = \mathscr{R}(u_i, v_j)$, each row vector $(r_{i1}, r_{i2}, \ldots, r_{in})$ represents a fuzzy set. The set of all fuzzy relations from $U$ to $V$, denoted as $\mathscr{F}(U \times V)$.

Let $\mathscr{R}$ be a fuzzy relation on $U$, i.e. $\mathscr{R} \in \mathscr{F}(U \times U)$. For $\forall x, y, z \in U$, we have

(1) $\mathscr{R}$ is reflexive $\Longleftrightarrow \mathscr{R}(x, x) = 1$;
(2) $\mathscr{R}$ is symmetric $\Longleftrightarrow \mathscr{R}(x, y) = \mathscr{R}(y, x)$;
(3) $\mathscr{R}$ is transitive $\Longleftrightarrow \mathscr{R}(x, z) \geqslant \vee_{y \in U}(\mathscr{R}(x, y) \wedge \mathscr{R}(y, z))$.

If $\mathscr{R}$ satisfies reflexivity and symmetry, then $\mathscr{R}$ is said to be a fuzzy similarity relation on $U$; if $\mathscr{R}$ satisfies reflexivity, symmetry, and transitivity, then $\mathscr{R}$ is said to be a fuzzy equivalence relation on $U$.

The union, intersection, and complement operations of the fuzzy set are extended to the general triangular norm ($t$-norm), triangular conorm ($s$-norm or $t$-conorm) and the negator operator [19].

Let the function $\mathscr{T} : [0, 1] \times [0, 1] \to [0, 1]$, if $\forall a, b, c \in [0, 1]$, it satisfies the following conditions:

(1) Commutativity: $\mathscr{T}(a, b) = \mathscr{T}(b, a)$;
(2) Associativity: $\mathscr{T}(\mathscr{T}(a, b), c) = \mathscr{T}(a, \mathscr{T}(b, c))$;
(3) Monotonicity: $b \leqslant c \Rightarrow \mathscr{T}(a, b) \leqslant \mathscr{T}(a, c)$;
(4) Boundary condition: $\mathscr{T}(a, 1) = a$,

then $\mathscr{T}$ is called the triangular norm on [0,1]. If the binary operator $\mathscr{S}$ satisfies the above commutativity, associativity, monotonicity, and boundary condition $\mathscr{S}(a, 0) = a$, then the binary operator $\mathscr{S}$ is called the triangle conorm on [0,1].

Negator operator $\mathscr{N}$ is a descending map on $[0, 1] \to [0, 1]$ that satisfies the boundary conditions $\mathscr{N}(0) = 1$ and $\mathscr{N}(1) = 0$. Usually, $\mathscr{N}(a) = 1 - a$ is called the standard negator operator.

Table 1 shows some triangular norms and conorms commonly used in fuzzy reasoning.

At present, the fuzzy rough set refers to the concept first proposed by Dubois and Prade [12], which is defined as follows.

**Definition 1.** Given $FIS = (U, A, V, f)$, let $\mathscr{R}$ be a fuzzy equivalence relation on $U$. For $\forall \mathscr{X} \in \mathscr{F}(U)$, the lower approximation $\underline{\mathscr{R}}\mathscr{X}$ and upper approximation $\overline{\mathscr{R}}\mathscr{X}$ of $\mathscr{X}$ are a pair of fuzzy sets on $U$ whose membership functions respectively are

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

$$\overline{\mathscr{R}}\mathscr{X}(x) = \sup_{y \in U} \min\{\mathscr{R}(x, y), \mathscr{X}(y)\}. \tag{2}$$

For the sake of exploring more complex relationships and structures between objects, Yeung et al. established a generalized fuzzy rough set model based on $t$-norm and $t$-conorm [40].

**Table 1**
Some common triangular norms and conorms.

| $\mathscr{T}(a,b)$ | $\mathscr{S}(a,b)$ |
|---|---|
| $\min\{a,b\}$ | $\max\{a,b\}$ |
| $ab$ | $a+b-ab$ |
| $\max\{a+b-1,0\}$ | $\min\{a+b,1\}$ |

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

$$\underline{\mathscr{R}}\mathscr{X}(x) = \inf_{y\in U} \mathscr{S}(\mathscr{N}(\mathscr{R}(x,y)), \mathscr{X}(y)), \tag{3}$$

$$\overline{\mathscr{R}}\mathscr{X}(x) = \sup_{y\in U} \mathscr{T}(\mathscr{R}(x,y), \mathscr{X}(y)). \tag{4}$$

We can discover that different FRS models can be obtained when different $\mathscr{S}$ and $\mathscr{T}$ are used. In practice, we can choose different $\mathscr{S}$ and $\mathscr{T}$ as needed. Clearly, when $\mathscr{T}(x,y) = \min\{x,y\}$ and $\mathscr{S}(x,y) = \max\{x,y\}$, the fuzzy rough set model in Definition 1 can be obtained.

## 3. Unsupervised mixed attribute reduction based on fuzzy rough sets

In this section, we first propose a generalized unsupervised mixed attribute reduction model, then analyze it based on the specific model. Further, we design a corresponding attribute reduction algorithm. Finally, an example is given to illustrate it.

### 3.1. Generalized model

**Definition 3.** Let $B = \{c_{j_1}, c_{j_2}, \cdots, c_{j_h}\}(1 \leqslant h \leqslant m) \subseteq C$, $\mathscr{R}_B$ is a fuzzy similarity relation induced by $B$ on $U$, then the fuzzy relation matrix $M(\mathscr{R}_B)$ is represented as

$$M(\mathscr{R}_B) = \begin{pmatrix} r_{11}^B & r_{12}^B & \cdots & r_{1n}^B \\ r_{21}^B & r_{22}^B & \cdots & r_{2n}^B \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}^B & r_{n2}^B & \cdots & r_{nn}^B \end{pmatrix} \tag{5}$$

A crisp equivalence relation produces a crisp partition, and a fuzzy similarity relation produces a general fuzzy partition.

**Definition 4.** Given $FIS = (U, C, V, f)$ and $\forall B \subset C$. The general fuzzy partition $U/\mathscr{R}_B$ of $U$ generated by $\mathscr{R}_B$ is defined as

$$U/\mathscr{R}_B = \left\{ [x_1]_{\mathscr{R}_B}, [x_2]_{\mathscr{R}_B}, \cdots, [x_n]_{\mathscr{R}_B} \right\}, \tag{6}$$

where $[x_i]_{\mathscr{R}_B} = \frac{r_{i1}^B}{x_1} + \frac{r_{i2}^B}{x_2} + \ldots + \frac{r_{in}^B}{x_n} = (r_{i1}^B, r_{i2}^B, \ldots, r_{in}^B)$ is called a fuzzy rough granule generated by $\mathscr{R}_B$.

Obviously, $[x_i]_{\mathscr{R}_B}$ is a fuzzy set on $\mathscr{R}_B$. We have $[x_i]_{\mathscr{R}_B}(x_j) = \mathscr{R}_B(x_i, x_j) = r_{ij}^B$. If $\mathscr{R}_B(x_i, x_j) = 1$, which means that $x_j$ must belong to $[x_i]_{\mathscr{R}_B}$; If $\mathscr{R}_B(x_i, x_j) = 0$, then $x_j$ definitely does not belong to $[x_i]_{\mathscr{R}_B}$. The cardinality of the fuzzy rough granule $[x_i]_{\mathscr{R}_B}$ is calculated by $|[x_i]_{\mathscr{R}_B}| = \sum_{j=1}^{n} \mathscr{R}_B(x_i, x_j)$. We can get $1 \leqslant |[x_i]_{\mathscr{R}_B}| \leqslant n$.

The common methods for the determination of $\mathscr{R}_B(x_i, x_j)$ are as follows [43]. (1) quantitative product method, (2) angle cosine method, (3) correlation coefficient method, (4) closeness method, (5) distance method, and (6) conjunction method. Among them, the conjunction method is used in most literatures [43], which is calculated by $\mathscr{R}_B(x_i, x_j) = \wedge_{l=1}^{h} \mathscr{R}_{c_{k_l}}(x_i, x_j)$. In this paper, the sixth method is adopted.

Obviously, the following properties hold.

**Property 1.** $B \subseteq C \Rightarrow \mathscr{R}_C \subseteq \mathscr{R}_B$.

**Property 2.** $B \subseteq C \Rightarrow [x_i]_{\mathscr{R}_C} \subseteq [x_i]_{\mathscr{R}_B}$.

In a FIS, different attribute subsets may induce different fuzzy similarity relation. Therefore, we may define the fuzzy lower and upper approximations considering the fuzzy partitions induced by different fuzzy similarity relations.

**Definition 5.** For $\forall P, Q \subset C$. Given $\forall [x_i]_{\mathscr{R}_Q} \in U/\mathscr{R}_Q$, the lower approximation $\underline{\mathscr{R}_P}[x_i]_{\mathscr{R}_Q}$ and upper approximation $\overline{\mathscr{R}_P}[x_i]_{\mathscr{R}_Q}$ of $[x_i]_{\mathscr{R}_Q}$ with respect to $\mathscr{R}_P$ are a pair of fuzzy sets on $U$ whose membership functions respectively are

$$\underline{\mathscr{R}_P}[x_i]_{\mathscr{R}_Q}(x) = \inf_{y\in U} \mathscr{S}\Big(\mathscr{N}(\mathscr{R}_P(x,y)), [x_i]_{\mathscr{R}_Q}(y)\Big), \tag{7}$$

$$\overline{\mathscr{R}_P}[x_i]_{\mathscr{R}_Q}(x) = \sup_{y\in U} \mathscr{T}\Big(\mathscr{R}_P(x,y), [x_i]_{\mathscr{R}_Q}(y)\Big). \tag{8}$$

In order to characterize the correlation between conditional attributes, we can define the dependency of the conditional attributes in fuzzy rough sets. Dependency is an important concept in the classical rough set theory, which analyzes the correlation between attributes through the perspective of the positive region. It has been widely used in supervised attribute reduction. Based on this, fuzzy dependency is defined to measure the correlation between the conditional attribute. To this end, we first give the following definition of a fuzzy positive region.

**Definition 6.** The fuzzy positive region of $Q$ relative to $P$ can be defined as

$$POS_P(Q) = \bigcup_{[x_i]_{\mathscr{R}_Q}\in U/\mathscr{I}} \underline{\mathscr{R}_P}[x_i]_{\mathscr{R}_Q}. \tag{9}$$

The fuzzy positive region is defined as the union of the lower approximations of $Q$ with respect to $P$. As the membership of the object in the fuzzy positive region increases, the likelihood of it being correctly classified into $U/\mathscr{R}_{\mathscr{I}}$ also increases. Obviously, those objects with larger membership are easier to classify into their respective fuzzy rough granules. In fact, the size of the aforementioned fuzzy positive region can reflect the correlation of conditional attributes.

An important issue in data analysis is finding correlations between conditional attributes. Under some circumstances, the correlation between conditional attributes may be partial, which means that $Q$ is only related to $P$ to a certain extent. Partial correlation can be defined by the fuzzy dependency.

**Definition 7.** If there is

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} = \frac{\sum_{x\in U} POS_P(Q)(x)}{|U|}, \tag{10}$$

we call $Q$ dependent on $P$ at $k$ degree, and write it as $P \Rightarrow_k Q$.

Obviously, $0 \leqslant k \leqslant 1$. It reflects the correlation between conditional attributes. When $k = 1$, we say that $Q$ completely depends on $P$; when $0 < k < 1$, we say that $Q$ depends in part on $P$; when $k = 0$, we say $Q$ is completely independent of $P$. Obviously, we should find a subset of attributes with the most fuzzy dependency, because in this case the attributes are more related.

**Property 3.** Given $Q \subseteq C$, for $\forall P_1 \subseteq P_2 \subseteq C$, we have

$$POS_{P_1}(Q) \subseteq POS_{P_2}(Q) and \gamma_{P_1}(Q) \leqslant \gamma_{P_2}(Q). \tag{11}$$

**Proof.** Given $Q \subseteq C$, for $\forall P_1 \subseteq P_2 \subseteq C$, according to Property 1, we have $\mathscr{R}_{P_2} \subseteq \mathscr{R}_{P_1}$. For $\forall y \in [x_i]_{\mathscr{R}_Q}$ and $x \in U$, we have $\mathscr{N}(\mathscr{R}_{P_2}(x,y)) = 1 - \mathscr{R}_{P_2}(x,y) \geqslant 1 - \mathscr{R}_{P_1}(x,y) = \mathscr{N}(\mathscr{R}_{P_1}(x,y))$. By the lower approximation in Definition 5 and the monotonicity of $\mathscr{S}$, we can get $\underline{\mathscr{R}_{P_1}}[x_i]_{\mathscr{R}_Q}(x) \leqslant \underline{\mathscr{R}_{P_2}}[x_i]_{\mathscr{R}_Q}(x)$. Therefore, $POS_{P_1}(Q) \subseteq POS_{P_2}(Q)$ and $\gamma_{P_1}(Q) \leqslant \gamma_{P_2}(Q)$.

**Property 4.** Given $P \subseteq C$, for $\forall Q_1 \subseteq Q_2 \subseteq C$, we have

$$POS_P(Q_2) \subseteq POS_P(Q_1) and \gamma_P(Q_2) \leqslant \gamma_P(Q_1). \tag{12}$$

**Proof.** Given $P \subseteq C$, for $\forall Q_1 \subseteq Q_2 \subseteq C$, according to Property 2, we have $[x_i]_{\mathscr{R}_{Q_2}} \subseteq [x_i]_{\mathscr{R}_{Q_1}}$. For $\forall y \in [x_i]_{\mathscr{R}_Q}$ and $x \in U$, we have $[x_i]_{\mathscr{R}_{Q_2}}(y) \leqslant [x_i]_{\mathscr{R}_{Q_1}}(y)$. By the lower approximation in Definition 5 and the monotonicity of $\mathscr{S}$, we can get $\underline{\mathscr{R}_P}[x_i]_{\mathscr{R}_{Q_2}}(x) \leqslant \underline{\mathscr{R}_P}[x_i]_{\mathscr{R}_{Q_1}}(x)$. Therefore, $POS_P(Q_2) \subseteq POS_P(Q_1)$ and $\gamma_P(Q_2) \leqslant \gamma_P(Q_1)$.

The above properties show the monotonicity of the fuzzy dependency function and attribute subset. These properties are crucial for designing a forward search algorithm because it guarantees that adding candidate attributes to the selected attribute subset will not reduce (or increase) the fuzzy dependency of the new attribute subset. Therefore, it is feasible to use fuzzy dependence as a criterion for unsupervised attribute reduction.

The fuzzy dependency $\gamma_P(Q)$ is used to measure the correlation between conditional attribute subsets $P$ and $Q$. In order to calculate the fuzzy dependency between the subsets of conditional attributes, we should not try to check all subsets of $C$. Since there are $2^{|C|}$ subsets of the attribute $C$, we will get $2^{2|C|}$ fuzzy dependencies. It is impractical to try to calculate the fuzzy dependencies of all subsets, because the time complexity will be exponential relative to $|C|$. Therefore, we only consider the fuzzy dependency between a partial subset of attributes, that is, the fuzzy dependency on all single attribute subsets.

Based on all single attribute subsets, the relevance is defined to measure the correlation of a conditional attribute subset.

**Definition 8.** The relevance of $B$ on all single attribute subsets is defined as

$$Rel_B\left(C^{\star}\right) = \frac{1}{m}\sum_{j=1}^{m}\left(\gamma_B\{c_j\}\right). \tag{13}$$

Obviously, $0 \leqslant Rel_B\left(C^{\star}\right) \leqslant 1$. The relevance reflects the correlation between a subset of conditional attributes and all sigle conditional attributes, which can be used as the definition of the significance of attributes.

**Property 5.** $B_1 \subseteq B_2 \subseteq C$, we have

$$Rel_{B_1}\left(C^{\star}\right) \leqslant Rel_{B_2}\left(C^{\star}\right). \tag{14}$$

**Definition 9.** If $Rel_B\left(C^{\star}\right) = Rel_{B-b}\left(C^{\star}\right)$, we say that $b$ in $B$ is redundant with respect to $C^{\star}$; otherwise, we say that $b$ in $B$ is indispensable relative to $C^{\star}$. If any one of $B$ is indispensable, we say that $B$ is independent.

**Definition 10.** Let $B \subseteq C$, we say that $B$ is a reduction of $C$, if $B$ satisfies

(1) $Rel_B\left(C^{\star}\right) = Rel_C\left(C^{\star}\right)$;
(2) $\forall b \in B, Rel_{B-b}\left(C^{\star}\right) < Rel_B\left(C^{\star}\right)$.

In the above definition, the first condition requires that a reduction cannot reduce the FIS's discrimination ability. The reduction should have the same resolution ability as all conditional attributes. The second condition requires that there are no extra attributes in the reduction, i.e., all attributes should be indispensable. This definition is completely consistent with the definition of supervised attribute reduction in classical rough sets. In the above definition, the fuzzy dependency is used to measure the correlation between conditional attributes. While, in the supervised attribute reduction, the decision attributes is used to define the significance of an attribute. Therefore, they are suitable for completely different application scenarios.

The relevance indicates the correlation between conditional attributes. Therefore, it can be used as an evaluation index of the significance of an attribute.

**Definition 11.** For $\forall b \in C - B$, the significance of $b$ in $B$ *with respect to* $C^{\star}$ *is defined* as

$$Sig\left(b, B, C^{\star}\right) = Rel_{B\cup\{b\}}\left(C^{\star}\right) - Rel_B\left(C^{\star}\right). \tag{15}$$

Since $0 \leqslant Rel_B\left(C^{\star}\right) \leqslant 1$ and $Rel_{B\cup\{b\}}\left(C^{\star}\right) \geqslant Rel_B\left(C^{\star}\right)$, we have $0 \leqslant Sig\left(b, B, C^{\star}\right) \leqslant 1$. If $Sig\left(b, B, C^{\star}\right) = 0$, we say that the attribute $b$ in $B$ is redundant with respect to $C^{\star}$, otherwise $b$ is indispensable.

### 3.2. A specific model

So far, we have established a generalized unsupervised mixed attribute reduction model called "unsupervised attribute reduction based on fuzzy rough sets (FRUAR)". The model construction route includes 5 parts: (1) fuzzy rough set model, (2) fuzzy positive region and dependency, (3) relevance, (4) significance for unsupervised attribute reduction. The FRUAR model has strong generalization ability, because different $\mathscr{T}$ and $\mathscr{S}$ can get different fuzzy rough set models, which can get different attribute reduction results. Next, a specific unsupervised mixed attribute reduction method based on fuzzy rough sets is analyzed.

The descriptions of objects often differ in magnitude and dimension. In order to obtain accurate data processing results and avoid the influence on different data, the original numerical attribute values need to be normalized first [44]. This paper uses the min–max normalization, which is calculated as follows.

$$F(f(x_i, c_k)) = \frac{f(x_i, c_k) - \min_{c_k}}{max_{c_k} - \min_{c_k}}, \tag{16}$$

where $\max_{c_k}$ and $\min_{c_k}$ are the maximum and minimum values of $c_k$, respectively. After normalization, the range of attribute values for these attributes is [0,1].

To effectively process nominal, numerical, and mixed attribute data, the fuzzy similarity $r_{ij}^{c_k}$ between objects $x_i$ and $x_j$ on $c_k$ is calculated as

$$r_{ij}^{c_k} = \begin{cases} 1, f(x_i, c_k) = f(x_j, c_k) \text{ and } c_k \text{ is nominal;} \\ 0, f(x_i, c_k) \neq f(x_j, c_k) \text{ and } c_k \text{ is nominal;} \\ 1 - |f(x_i, c_k) - f(x_j, c_k)|, |f(x_i, c_k) - f(x_j, c_k)| \leqslant \varepsilon_k \text{and} c_k \text{ is numerical;} \\ 0, |f(x_i, c_k) - f(x_j, c_k)| > \varepsilon_k \text{ and } c_k \text{ is numerical,} \end{cases} \tag{17}$$

Obviously, $r_{ij}^{c_k} = r_{ji}^{c_k}$ and $r_{ii}^{c_k} = 0, 0 \leqslant r_{ij}^{c_k} \leqslant 1, M_{\mathcal{R}_{\varepsilon_k}}$ is a fuzzy similarity relation. $\varepsilon_{c_k}$ is an adjustable fuzzy similarity radius, whch is calculated as follows

$$\varepsilon_{c_k} = \frac{std(c_k)}{\lambda}, \tag{18}$$

where $std(c_k)$ is the standard deviation of the attribute value of $c_k$, and the preset parameter $\lambda$ is used to adjust the fuzzy similarity radius.

**Algorithm 1:** The algorithm FRUAR

**Input**: $FIS = (U, C, V, f)$, threshold $\lambda$, $|C| = m$
**Output**: A Reduction (R)

$R \leftarrow \emptyset$, Label$\leftarrow$ 1, $B \leftarrow C - R$;
**for** $k \leftarrow 1$ *to* $m$ **do**
  Calculate $M_{\mathcal{R}_{c_k}}$;
**end**
**while** Label **do**
  //Let $B = \{c_{k_1}, c_{k_2}, \cdots, c_{k_h}\}$
  **for** $l \leftarrow 1$ *to* $h$ **do**
    Calculate $M_{\mathcal{R}_{R\cup\{c_{k_l}\}}}$;
    **for** $s \leftarrow 1$ *to* $m$ **do**
      Calculate $\gamma_{R\cup\{c_{k_l}\}}(\{c_s\})$;
    **end**
    Calculate $Rel_{R\cup\{c_{k_l}\}}(C^\star)$;
  **end**
  Select attribute $c_{k_{l'}}$, s.t. $Rel_{R\cup\{k_{l'}\}}(C^\star)$ have a maximum value;
  Calculate $Sig(c_{k_{l'}}, R, C^\star) = Rel_{R\cup c_{k_{l'}}}(C^\star) - Rel_R(C^\star)$;
  **if** $Sig(c_{k_{l'}}, R, C^\star) > 0$ **then**
    $R \leftarrow R \cup \{c_{k_{l'}}\}$, $B \leftarrow B - R$;
    **else**
      Label$\leftarrow$ 0;
    **end**
  **end**
**end**
**if** $|R| == m$ **then**
  **return** $R \leftarrow R(1 : m - 1)$;
  **else**
    **return** $R$.
  **end**
**end**

Next, we set $\mathscr{T}(x, y) = \min\{x, y\}, \mathscr{S}(x, y) = \max\{x, y\}$. When the relationship between objects is a crisp equivalence relation and the subset of objects to be approximated is a fuzzy set, the model will be downgraded to a rough fuzzy set model; if the subset of objects to be approximated is crisp and the relationship between objects is a fuzzy relation, the model is a fuzzy rough set model; in addition, when both are equivalence relations, it is a rough set model. These characteristics make the model suitable for attribute reduction of mixed attribute data.

### 3.3. A specific algorithm

In this section, we design a fuzzy rough sets-based unsupervised attribute reduction (FRUAR) algorithm and analyze its complexity.

Algorithm 1 starts with the empty set and calculates the significance of all remaining attributes each time. Thus the attribute with the highest significance is selected and added to the reduction set until the significance of all remaining attributes is 0.

In Algorithm 1, the number of loops in Steps 2-4 is $m$, and the number of loops in Step 3 is $n$. In addition, the number of loops in Steps 7-13 is $h$, the number of loops in Steps 9-11 is $m$, and the number of loops in Step 8 is $n$. Thus, the total number of loops of Algorithm 1 is $m \times n + h \times m \times n$. Therefore, in the worst case, the time complexity of Algorithm 1 is $O(m^2 n)$.

### 3.4. A specific example

This section illustrates the FRUAR method through an example.

**Example 1.** A $FIS = (U, V, C, f)$ is on the left side of Table 2. It mainly relates to mixed attribute data. Herein, $U = \{x_1, x_2, \cdots, x_6\}, C = \{c_1, c_2, c_3, c_4\}$. Among them, $c_1, c_2$ are nominal attributes, and $c_2, c_3$ are numeric attributes.

To begin with, the min–max normalization (Eq. (16)) is performed on the raw numerical data, and the normalized result is shown on the right side of Table 2. The standard deviations of numeric attributes $c_3$ and $c_4$ are $std(c_3) \approx 0.3492, std(c_4) \approx 0.3146$, respectively. According to the Eq. (18) and let $\lambda = 1$, the fuzzy similarity radii are calculated as $std(c_3) \approx 0.3492, std(c_4) \approx 0.3146$. Suppose $R = \varnothing, B = C$.

For any $c_k \in C$, the fuzzy similarity relation matrices are as follows.

$$M(\mathscr{R}_{c_1}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}, M(\mathscr{R}_{c_2}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$M(\mathscr{R}_{c_3}) = \begin{pmatrix} 1 & 0 & 0.7500 & 0 & 0 & 0.8750 \\ 0 & 1 & 0.7500 & 0.8750 & 0 & 0 \\ 0.7500 & 0.7500 & 1 & 0 & 0 & 0 \\ 0 & 0.8750 & 0 & 1 & 0.7500 & 0 \\ 0 & 0 & 0 & 0.7500 & 1 & 0 \\ 0.8750 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$M(\mathscr{R}_{c_4}) = \begin{pmatrix} 1 & 1 & 0.8750 & 0 & 0 & 0 \\ 1 & 1 & 0.8750 & 0 & 0 & 0 \\ 0.8750 & 0.8750 & 1 & 0.7500 & 0 & 0 \\ 0 & 0 & 0.7500 & 1 & 0.7500 & 0 \\ 0 & 0 & 0 & 0.7500 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

By using Eq. 7, the dependencies between two sing attributes are calculated as follows

$\gamma_{\{c_1\}}(\{c_1\}) = \frac{\sum_{x \in U} POS_{\{c_1\}}(\{c_1\})}{|U|} = 1, \gamma_{\{c_1\}}(\{c_2\}) \approx 0.1667, \gamma_{\{c_1\}}(\{c_3\}) \approx 0.4167, \gamma_{\{c_1\}}(\{c_4\}) \approx 0.4167;$

$\gamma_{\{c_2\}}(\{c_1\}) = \frac{\sum_{x \in U} POS_{\{c_2\}}(\{c_1\})}{|U|} \approx 0.6667, \gamma_{\{c_2\}}(\{c_2\}) = 1, \gamma_{\{c_2\}}(\{c_3\}) \approx 0.5883, \gamma_{\{c_2\}}(\{c_4\}) \approx 0.5883;$

$\gamma_{\{c_3\}}(\{c_1\}) = \frac{\sum_{x \in U} POS_{\{c_3\}}(\{c_1\})}{|U|} \approx 0.2917, \gamma_{\{c_3\}}(\{c_2\}) \approx 0.1667, \gamma_{\{c_3\}}(\{c_3\}) \approx 0.7708, \gamma_{\{c_3\}}(\{c_4\}) \approx 0.4792;$

$\gamma_{\{c_4\}}(\{c_1\}) = \frac{\sum_{x \in U} POS_{\{c_4\}}(\{c_1\})}{|U|} \approx 0.3958, \gamma_{\{c_4\}}(\{c_2\}) \approx 0.2708, \gamma_{\{c_4\}}(\{c_3\}) \approx 0.6250, \gamma_{\{c_4\}}(\{c_4\}) \approx 0.8333.$

Further, the relevance of $c_k \in C$ on each attribute is calculated as follows.

$Rel_{c_1}\left(C^{\Leftrightarrow}\right) = \frac{1}{m}\sum_{k=1}^{m} \gamma_{c_1}(\{c_k\}) = \frac{1}{4}(1 + 0.1667 + 0.4167 + 0.4167) \approx 0.5000$; $Rel_{c_2}\left(C^{\Leftrightarrow}\right) \approx 0.7083$;

$Rel_{c_3}\left(C^{\Leftrightarrow}\right) \approx 0.4271$; $Rel_{c_4}\left(C^{\Leftrightarrow}\right) \approx 0.5313$.

Accordingly, the significance of each attribute is calculated as follows.

$Sig\left(c_1, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_1}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 0.5000;$

$Sig\left(c_2, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_2}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 0.7083;$

$Sig\left(c_3, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_3}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 0.4271;$

$Sig\left(c_4, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_4}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 0.5313.$ ($R = \varnothing$, so $Rel_R\left(C^{\Leftrightarrow}\right) = 0$).

Next, the attribute with the highest significance is selected to add to $R$, that is, the attribute $c_2$ is added to $R$, and we can get $R = \{c_2\}, B = C - \{c_2\} = \{c_1, c_3, c_4\}$.

Similarly, we can calculate

$Sig\left(c_1, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_1}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 0.7917 - 0.7083 = 0.1563;$

$Sig\left(c_3, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_3}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 1.0000 - 0.7083 = 0.3646;$

$Sig\left(c_4, R, C^{\Leftrightarrow}\right) = Rel_{R \cup c_4}\left(C^{\Leftrightarrow}\right) - Rel_R\left(C^{\Leftrightarrow}\right) = 0.8750 - 0.7083 = 0.2396.$

**Table 2**
Initial and normalized FIS.

| $U$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | $b$ | D | 10 | 0.7 | $b$ | D | 0.8750 | 0.6250 |
| $x_2$ | $c$ | A | 6 | 0.3 | $c$ | A | 0.3750 | 0.6250 |
| $x_3$ | $c$ | B | 2 | 0.5 | $c$ | B | 0.6250 | 0.5000 |
| $x_4$ | $a$ | B | 3 | 0.2 | $a$ | B | 0.2500 | 0.2500 |
| $x_5$ | $a$ | C | 7 | 0.4 | $a$ | C | 0 | 0 |
| $x_6$ | $c$ | A | 3 | 0.6 | $c$ | A | 1 | 1 |

At this point, we can get $R = \{c_2, c_3\}, B = \{c_1, c_4\}$.

Further, we can calculate

$$Sig\left(c_1, R, C^{\star}\right) = Rel_{R\cup c_1}\left(C^{\star}\right) - Rel_R\left(C^{\star}\right) = 1 - 1 = 0;$$

$$Sig\left(c_4, R, C^{\star}\right) = Rel_{R\cup c_4}\left(C^{\star}\right) - Rel_R\left(C^{\star}\right) = 1 - 1 = 0.$$

Therefore, $R = \{c_2, c_3\}$.

## 4. Experiments and analyses

This section evaluates the algorithm FRUAR through the performance in classification, clustering, and outlier detection tasks. To this end, thirty data sets (including numerical, nominal, and mixed attributes) are selected from the UCI database for experiments [10]. For missing values in some data sets, this paper uses the maximum probability value method to fill the missing values, that is, to fill the missing attribute values with the value that appears most frequently on other objects on this attribute. The description of data sets used for classification and clustering is shown in Table 3.

On the data sets listed in Table 3, we compared the algorithm FRUAR with Laplacian Score-based (LS) [17], Feature Similarity-based Feature Selection (FSFS) [26], SPECtral analysis-based (SPEC) [48], UnSupervised Quick Reduct (USQR) [32], Unsupervised Entropy Based Reduct (UEBR) [33], and Unsupervised Fuzzy Rough based Feature Selection (UFRFS) [24]. Among them, algorithms LS, FSFS, and SPEC are usually applied to numerical attribute data. Both algorithms USQR and UEBR are attribute reduction algorithms based on rough sets. They are only applicable to nominal attribute data, while numerical attribute data requires discretization preprocessing. Although the algorithm UFRFS is an unsupervised feature selection based on fuzzy rough sets, it does not consider mixed attributes.

The selected data sets are mostly used for the evaluation of classification and clustering methods. However, few data sets exist for the evaluation of outlier detection. Therefore, this article uses the downsampling method proposed by [4] to obtain the outlier detection data set. Following the preprocessing method in [42], The related description of data sets for outlier detection is shown in Table 4.[1]

### 4.1. Experimental preparation

For the classification experiments, we call Classification And Regression Trees (CART), Naive Bayes (NB), and k-Nearest Neighbor (kNN) classification algorithm in Matlab R2015b to evaluate the performance of these comparison algorithms. All classification experiments are performed by 10-fold cross-validation. In addition, the "DistributionNames" of the algorithm NB is set to "kernel". we repeat 10 times experiments, and the average and standard deviation of the classification accuracy are calculated as the final results.

For the clustering experiments, the performance of these comparison algorithms is also evaluated by calling the algorithm k-Means in Matlab R2015b. The number of clusters is set to the number of real decision classes. The clustering algorithm returns the predicted decisions of the data set. We use clustering accuracy (ACC) to evaluate the clustering performance [50]. The larger the ACC of a clustering algorithm is, the better performance it will be. Since these clustering algorithms have randomness, we also repeat them 10 times experiments. The average and standard deviation of the clustering accuracy are calculated as the final results.

In addition, for the outlier experiments, Distance-based (DIS) [22], k-Nearest Neighbor Outlier (kNNO) [29], and density-based (Local outlier factor, LOF) [3] algorithms are used to test the outlier detection effect of the comparison algorithms.

In general, Precision (P), Recall (R), True Positive Rate (TPR), False Positive Rate (FPR), False Negative Rate (FNR), and Receiver Operating Characteristic (ROC) curves can be used to evaluate the effectiveness of the proposed method [1,42]. The specific steps are as follows.

In the outlier detection, most of the detection methods ultimately output the outlier factor of each object in $U$, and the larger the outlier factor of an object, the more likely it is the outlier. These objects can be arranged in descending order according to their outlier factor values. Supposing an order number $t$, objects with a sequence number greater than or equal

[1] https://github.com/BElloney/Outlier-detection.

**Table 3**
The description of data sets for classification and clustering.

| ID | Data set | Object | Conditional attribute | Decision class | Type |
|---|---|---|---|---|---|
| 1 | Abalone | 4177 | 8 | 29 | Mixed |
| 2 | Anneal | 798 | 38 | 6 | Mixed |
| 3 | Autos | 205 | 25 | 6 | Mixed |
| 4 | Bands | 531 | 39 | 2 | Mixed |
| 5 | Credit approval | 690 | 15 | 2 | Mixed |
| 6 | German | 1000 | 20 | 2 | Mixed |
| 7 | Heart disease | 270 | 13 | 2 | Mixed |
| 8 | Sick | 3772 | 29 | 2 | Mixed |
| 9 | Dermatology | 366 | 34 | 6 | Numeric |
| 10 | Ionosphere | 351 | 33 | 2 | Numeric |
| 11 | Movement_libras | 360 | 90 | 15 | Numeric |
| 12 | Page_blocks | 5473 | 10 | 2 | Numeric |
| 13 | Parkinsons | 768 | 22 | 2 | Numeric |
| 14 | Sonar | 208 | 60 | 2 | Numeric |
| 15 | WDBC | 569 | 31 | 2 | Numeric |
| 16 | Winequality_red | 1599 | 11 | 6 | Numeric |
| 17 | Winequality_white | 4898 | 11 | 7 | Numeric |
| 18 | Yeast | 1484 | 8 | 10 | Numeric |
| 19 | Chess | 3196 | 36 | 2 | Nominal |
| 20 | Lymphography | 148 | 18 | 4 | Nominal |
| 21 | Mushroom | 8124 | 22 | 2 | Nominal |

**Table 4**
The description of data sets for outlier detection.

| ID | Data set | Attribute | Object | Outlier | Type |
|---|---|---|---|---|---|
| 1 | Cred | 15 | 425 | 42 | Mixed |
| 2 | Heart | 13 | 166 | 16 | Mixed |
| 3 | Hepa | 19 | 94 | 9 | Mixed |
| 4 | Diab | 8 | 526 | 26 | Numeric |
| 5 | Iono | 34 | 249 | 24 | Numeric |
| 6 | Pima | 9 | 555 | 55 | Numeric |
| 7 | Sonar | 60 | 107 | 10 | Numeric |
| 8 | Wbc | 9 | 483 | 39 | Numeric |
| 9 | Lymp | 18 | 148 | 6 | Nominal |

to $t$ are treated as outliers. If the given $t$ is too small, it will cause the method to miss the true outliers. Conversely, too many objects are judged to be outliers, which leads to too excessive false positives. This trade-off can usually be measured by P and R. For $\forall t, OS(t)$ is a function of $t$, which denotes the outlier set detected by the given $t$. $OS^{\circ}$ represents the true outlier set in data set. $P(t)$ and $R(t)$ are defined as, respectively

$$P(t) = \frac{|OS(t) \cap OS^{\circ}|}{|OS(t)|} \times 100\%; \tag{19}$$

$$R(t) = \frac{|OS(t) \cap OS^{\circ}|}{|OS^{\circ}|} \times 100\%. \tag{20}$$

where $P(t)$ indicates the ratio of true outliers detected at a given $t$ to all outliers detected at a given $t$, and $R(t)$ represents the ratio of true outliers detected at a given $t$ to all true outliers.

$TPR(t), FPR(t)$,and $FNR(t)$ are calculated by, respectively

$$TPR(t) = R(t) = \frac{|OS(t) \cap OS^{\circ}|}{|OS^{\circ}|} \times 100\%; \tag{21}$$

$$FPR(t) = \frac{|OS(t) - OS^{\circ}|}{|U - OS^{\circ}|} \times 100\%; \tag{22}$$

$$FNR(t) = \frac{|OS^{\circ} \cap (U - OS(t))|}{|OS^{\circ}|} \times 100\%. \tag{23}$$

Under a given t, the larger the values of $P(t), R(t)$, and $TPR(t)$, the better the result of outlier detection. However, the smaller the value of $FPR(t)$ and $FNR(t)$, the better the result of outlier detection. What's more, it can be shown that $P(t)$ and $R(t)$ are equal and $P(t) + FNR(t) = 1$ when $t = |OS^{\circ}|$.

ROC is a curve with the False Positive Rate (FPR) as the abscissa and the True Positive Rate (TPR) as the ordinate. ROC curve can be used to compare the performance of different outlier detection models. If the curve of a detection method is closer to the upper left corner of the first quadrant, and the larger the area under the curve, the better the performance.

we repeat algorithms kNNO and LOF for different reduced attributes, and calculate the optimal detection results of k and MinPts in the range of $[1, n/4]$ with step size 1.

In the experiments, for algorithms LS, FSFS, and SPEC, all different nominal attribute values are replaced with different integer values, and all attribute values are normalized using the min–max normalization (Eq. (16)) into [0,1] interval. Following the preferred experimental setting in [50], the neighborhood size is fixed to be 5 when the algorithm LS calculates the neighbor graph matrix. For the algorithm FSFS, the feature similarity is calculated by the "Maximal Information Compression Index". For the algorithm FSFS, its style is set as $-1$. Generally, algorithms LS, FSFS, and SPEC output a sequence of attributes arranged in descending order of attribute importance. In order to compare with the FRUAR algorithm, we select attribute subsequences with the same number of attributes selected by the algorithm FRUAR. Because algorithms USRR and UEBR only consider nominal attributes, they need to use discretization algorithms such as equal width and equal frequency to pre-process the numerical attributes. In general, different discretization methods may result in different attribute subsets being selected. Experimental results in [20] show that the FCM discretization method provides better performance than equal width and equal frequency. Therefore, we use FCM technology to discretize the numerical atrributes [41]. Numerical attributes are discretized into 4 intervals. Different learning algorithms make use of available attributes differently. In other words, different learning algorithms may require different subsets of attributes to produce the best performance. Therefore, for the algorithm FRUAR, we calculate the optimal result of $\lambda$ in the range of [0.1,3] with step size 0.1. Besides, through a lot of experiments, we find that it is usually difficult to reach the stop condition that the significance of all remaining attributes is 0 for data sets containing numerical attributes in Algorithm 1. Therefore, the significance stop condition for data sets containing numerical attributes is set to 0.001.

### 4.2. Classification experiments

First, let's analyze the results of the classification experiments. The comparison results of classification experiment are shown in Tables 5–9. Among them, Table 5 gives the attributes and numbers selected by algorithms USQR, UEBR, and UFRFS.

**Table 5**
The attributes and number selected by algorithms USQR, UEBR, and UFRFS.

| Data set | Raw attribute | USQR(Number) | UEBR(Number) | UFRFS(Number) |
|---|---|---|---|---|
| Abalone | 8 | 2,1,3,4,6,8,7,5(8) | 5,3,1,4,7,8,6,2(8) | 1,2,3,4,5,6,7,8(8) |
| Anneal | 38 | 3,33,35,12,34,5,36,4,7,32, 1,37(12) | 3,35,33,34,12,8,5,4,36,37,32(11) | 1,3,4,5,7,8,9,12,13,17, 27,32,33,34,35,36,37,38(18) |
| Autos | 25 | 2,20,10,6,23,1,25,13,21,5,4 (11) | 2,21,6,22,12,24,5,25,9,4,13,23(12) | 1,2,4,5,6,7,11,13,17,18, 19,21,22,23,24,25(16) |
| Bands | 39 | 2,22,38,24 (4) | 2,22,38,24,27 (5) | 3,17,19,20,22,23,24,25,29,30, 31,32,34,37,38,39(16) |
| Credit approval | 15 | 4,10,6,2,14,3,8,1,12,9,7,13,11(13) | 6,2,3,14,8,10,4,1,12,7,9,13 (12) | 1,2,3,5,6,7,8,9,10,11,12,13,14,15 (14) |
| German | 20 | 4,3,7,12,2,8,1,13,11,6,18 (11) | 4,7,2,12,8,13,1,3,11,6,5 (11) | 1,2,3,4,5,6,7,8,9,10,11,12,13,14, 15,16,17,18,19,20(20) |
| Heart disease | 13 | 5,10,3,1,8,4,12,7,11,9 (10) | 8,1,10,4,5,3,12,7,11,9 (10) | 1,2,3,4,5,6,7,8,9,10,11,12,13 (13) |
| Sick | 29 | 24,29,1,26,20,2,22,19,23,3, 11,10,17,6,14,16,18,9,13,5(20) | 22,24,1,29,20,26,2,19,23,3, 11,10,17,6,16,14,18,9(18) | 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15, 16,17,18,19,20,21,22,23,24,25,26,29 (27) |
| Dermatology | 34 | 33,34,16,4,3,18,19,2,17,32,14 (11) | 20,33,34,3,16,4,19,32,18,14, 2,8,22(13) | 1,4,5,11,14,16,17,18,19,20,21,22,23, 26,28,31,32,33,3(19) |
| Ionosphere | 33 | 26,20,21,29,17,23,7,4,31,10, 2,33,24,12,5,32,25,27(18) | 14,27,28,19,31,7,30,4,33,17, 29,24,2,20,15,10,12,26(18) | 1,6,19,22,23,25,26,27,28,31,32,33 (12) |
| Movement_libras | 90 | 2,3,30,66,31,59,71,56,80,49, 40,85,37(13) | 4,11,76,55,33,50,47,40,77,41, 62,71,89(13) | 81,84,88,89,90(5) |
| Page_blocks | 10 | 4,2,1,5,6,10,9,7,3,8 (10) | 2,5,6,10,4,3,7,9,1,8 (10) | 1,2,4,5,6,7,9,10 (8) |
| Parkinsons | 22 | 16,18,20,2,22,11,21,17,5,3 (10) | 12,22,1,18,20,17,21,5,11,2,13 (11) | 2,3,14,17,18,19,20,21 (8) |
| Sonar | 60 | 17,13,21,26,37,51,7,36,9,54 (10) | 16,36,22,26,57,51,5,4,60 (9) | 50,53,54,55,57,58,59,60 (8) |
| WDBC | 31 | 4,19,11,3,30,6,13,16,28,5,12,10,18 (13) | 24,7,23,26,16,30,6,31,12,2,29,20,10 (13) | 1,20,21,23,25,26,28,29,30 (9) |
| Winequality_red | 11 | 1,5,8,6,2,11,10,9,3,4,7 (11) | 1,11,6,2,10,9,4,5,3,7,8 (11) | 1,2,3,4,5,6,7,9,10,11 (10) |
| Winequality_white | 11 | 1,5,4,11,9,7,10,2,6,3,8 (11) | 8,9,6,10,1,2,3,7,5,11,4 (11) | 1,2,3,4,5,6,7,9,10,11 (10) |
| Yeast | 8 | 4,7,1,3,2,8,6 (7) | 1,3,4,7,2,8,6 (7) | 1,2,3,4,5,6,7,8 (8) |
| Chess | 36 | 35,11,9,7,34,10,5,6,13,24, 33,15,36,18,26,12,20,21,4,23, 22,30,16,31,1,27(26) | 11,9,34,35,7,10,13,24,33,6, 5,15,36,26,18,12,20,21,4,22, 23,30,16,31,1,27(26) | 1,2,3,4,5,6,7,8,10,11,12,13,15,16, 17,18,19,20,21,22,23,24,25,26,27, 28,30,31,33,34,35,36 (32) |
| Lymphography | 18 | 18,14,12,1,15,2,16,8,11,3 (10) | 18,14,12,15,1,2,16,8,11,3 (10) | 1,2,3,5,6,7,8,10,11,12,13,14,15,16, 17,18(16) |
| Mushroom | 22 | 14,9,5,22,3,1,2,21,15,20, 13,12,17,7(14) | 9,5,22,3,1,2,15,14,21,20, 13,12,17,7(14) | 1,2,3,4,5,6,7,8,9,10,12,13,14,15,17, 20,21,22(18) |
| Average | 27.3 | 12.5 | 12.5 | 14.4 |

**Table 6**
The optimal attribute subsets and number of classification algorithms.

| Data set | Raw attribute | CART(Number) | $\lambda$ | NB(Number) | $\lambda$ | kNN(Number) | $\lambda$ |
|---|---|---|---|---|---|---|---|
| Abalone | 8 | 5,1,3,4,8,7,6(7) | 0.8 | 1,4,5,8,6,2 (6) | 0.1 | 5,1,4,8,2,6,7 (7) | 0.4 |
| Anneal | 38 | 3,33,12,32,8,34,35,5,36,9, 37,1,17(13) | 1 | 3,12,32,8,33,34,5,35,9,36, 37,1,17(13) | 0.7 | 3,33,12,34,8,32,35,5,36,37, 9,1,4(13) | 1.8 |
| Autos | 25 | 2,24,6,19,5,4,21,10,12,22,9(11) | 1.5 | 2,23,6,22,5,4,10,19,11,17, 1,7(12) | 1.4 | 2,23,6,22,12,21,5,9,19(9) | 2.2 |
| Bands | 39 | 2,22,24,29 (4) | 2.7 | 2,22,24,29 (4) | 2.7 | 2,22,24,29 (4) | 2.7 |
| Credit approval | 15 | 6,4,10,7,12,9,1,13(8) | 0.2 | 4,6,10,12,7,9,1,13(8) | 0.1 | 6,4,10,7,12,9,1,13,8(9) | 0.3 |
| German | 20 | 4,3,7,12,1,6,9,2,17,8,11,16(12) | 1.1 | 4,3,7,12,8,1,11,13,9,6,2(11) | 1.7 | 4,3,12,7,8,1,11,9,6,17,2 (11) | 1.2 |
| Heart disease | 13 | 3,13,11,7,9,2,6,12(8) | 0.1 | 3,13,11,7,9,2,6,12(8) | 0.1 | 8,5,4,1,3,12,10,11(8) | 3 |
| Sick | 29 | 26,29,24,1,20,2,19,23,3,11, 10,17,6,22,14,16,9,8,13,5,4,18 (22) | 1.6 | 26,29,24,1,20,2,19,23,3, 11,10,17,6,22,14,16,9,8, 13,5,4,18(22) | 1.6 | 29,23,2,19,3,11,10,17,26,14, 6,4,16,8,13,24,5,9,18,7,20, 12,21,1(24) | 0.3 |
| Dermatology | 34 | 27,28,4,20,30,11,13,8,15, 32,10,5,18,26,14,9,34,23, 3,24,19,17,1,16,21,31,2(27) | 0.5 | 21,33,11,28,4,9,15,16,5, 26,7,18,32,24,14,8,6,10, 22,3,17,29,30,13,20,34(26) | 0.3 | 33,21,11,28,4,15,9,16,5,26, 19,24,27,10,7,32,14,6,22,3, 13,29,20,8,17(25) | 0.1 |
| Ionosphere | 33 | 26,18,23,17,33,5,2,32, 15,9,6,3,4(13) | 3 | 26,18,23,17,33,5,2,32,15, 9,6,3,4(13) | 3 | 26,18,27,7,3,22,33,11,2,9, 32,6,17,23,25,19,12(17) | 2.4 |
| Movement_libras | 90 | 21,53,6,58,79,74,32,90,1,46, 35,63,89,20,64,82,36,41,71,15, 27,54,26,49,40(25) | 0.5 | 6,1,54,61,29,72,36,85,84,41, 51,42,19,62,90,20,75,46,30 (19) | 1.0 | 4,7,61,58,29,90,81,26(8) | 3.0 |
| Page_blocks | 10 | 9,5,6,2,1,10,4,7(8) | 2.6 | 9,5,6,2,4,1,10,7(8) | 1.3 | 9,5,6,2,4,1,10,7,8(9) | 1.4 |
| Parkinsons | 22 | 14,18,19,21,1,20,17,3,2(9) | 2.2 | 16,18,21,20,3,17,22,2(8) | 2.5 | 4,1,20,12,2,21,17,3,18,5, 19,11,15,16,13,7,22(17) | 0.5 |
| Sonar | 60 | 18,26,11,32,37,43,5,29,54, 14,53,34,47,60,21,58,50,7, 56,6,49,31,13,23,2(25) | 0.8 | 46,20,1,23,17,37,26,32,28, 55,10,35,40,8,60,50,45,21, 30,19,54,6,25,5,42,34,12, 18,29,27(30) | 0.1 | 1,46,20,23,17,37,59,3,58,10, 51,26,40,34,6,29,45,55,8,27, 32,60,19,25,35,2,57,52,21,50, 56,42,12,9,49,18,54,44,30 (39) | 0.3 |
| WDBC | 31 | 8,17,11,1,20,29,25,13,27, 12,3,18,10,21,26,16,31,30, 5,19,6,23,14(23) | 0.3 | 24,17,26,30,3,19,11,13,10, 20,9,12,16,31,1,6,23,5,28, 29,25,27,21,2(24) | 0.7 | 8,1,5,27,11,29,23,28,15,13, 21,22,20,26,9,3,17,10(18) | 0.1 |
| Winequality_red | 11 | 5,9,8,7,2,4,10,6,11,1(10) | 0.6 | 9,4,8,5,2,10,6,11,1,7(10) | 0.4 | 5,4,7,10,1,2,6,11,8,3(10) | 0.2 |
| Winequality_white | 11 | 8,5,3,6,2,1,10,9,7,11(10) | 0.7 | 8,6,5,3,1,2,11,10,9(9) | 0.1 | 5,3,2,1,4,7,9,10,6,11(10) | 1.7 |
| Yeast | 8 | 7,8,2,3,4,1,6(7) | 1 | 7,3,2,4,1,8,6(7) | 1.1 | 7,8,2,3,4,1,6 (7) | 1 |
| Chess | 36 | 35,11,9,7,34,10,5,6,13, 24,33,15,36,18,26,12,20, 21,4,23,22,30,16,31,1,27, 17,2,8,3,19,25,28(33) | – | 35,11,9,7,34,10,5,6,13,24, 33,15,36,18,26,12,20,21,4, 23,22,30,16,31,1,27,17,2, 8,3,19,25,28(33) | – | 35,11,9,7,34,10,5,6,13,24, 33,15,36,18,26,12,20,21,4, 23,22,30,16,31,1,27,17,2,8, 3,19,25,28(33) | – |
| Lymphography | 18 | 18,14,12,1,15,2,16,8,11,3(10) | – | 18,14,12,1,15,2,16,8,11,3 (10) | – | 18,14,12,1,15,2,16,8,11,3(10) | – |
| Mushroom | 22 | 14,9,5,22,3,1,2,21,15,20, 13,12,17,7,6(15) | – | 14,9,5,22,3,1,2,21,15,20, 13,12,17,7,6(15) | – | 14,9,5,22,3,1,2,21,15,20, 13,12,17,7,6(15) | – |
| Average | 27.3 | 14.3 | – | 14.1 | – | 14.4 | – |

Table 6 presents the optimal attribute subset and number of the algorithm FRUAR with respect to different classification algorithms. Tables 7–9 give the comparison of classification accuracy based on algorithms CART, NB, and kNN, respectively. The optimal values of different algorithms are in bold. "–" means that no reduction result is obtained.

From Tables 5 and 6, we can find that, in most cases, all attribute reduction algorithms can remove some candidate attributes. However, we can see from Table 5 that algorithms USQR, UEBR, and UFRFS cannot remove any candidate attributes on some data sets. For example, algorithms USQR and UEBR cannot remove any candidate attributes on Abalone, Page_blocks, Winequality_red, and Winequality_white data sets. The algorithm UFRFS cannot remove any candidate attributes on German and Yeast data sets. In addition, the average size of the attribute subsets selected using algorithms USQR and UEBR is slightly smaller than the average size of the attribute subsets selected using other algorithms. The reason for this result may be the loss of information caused by data discretization, resulting in lower classification accuracy. Besides, different classification algorithms may require attribute feature subsets to produce the best performance. For example, in Table 6, the optimal attribute subsets of the algorithm FRUAR based on CART, NB, and KNN on Abalone are 5,1,3,4,8,7,6(7), 1,4,5,8,6,2(6), and 5,1,4,8,2,6,7(7), respectively.

It can be seen from Tables 7–9 that the algorithm FRUAR achieves better classification accuracy. The specific analyses are as follows.

Table 7–9 show that the algorithm FRUAR has 14, 15, and 12 cases to achieve the optimal classification accuracy, respectively. For other algorithms, only relatively small cases achieve the best classification accuracy. For example, according to the results in Table 7, the algorithm FRUAR has 19 cases higher than algorithms LS and UEBR, 20 cases higher than algorithms FSFS, SPEC, and USQR, and 17 cases higher than the algorithm UFRFS, respectively.

**Table 7**
The comparison of classification accuracy of the algorithm CART on reduced data (%).

| Data set | Raw data | LS | FSFS | SPEC | USQR | UEBR | UFRFS | FRUAR |
|---|---|---|---|---|---|---|---|---|
| Abalone | 20.95 ± 0.42 | 21.06 ± 0.66 | 20.75 ± 0.60 | **21.41 ± 0.19** | – | – | – | 21.03 ± 0.66 |
| Anneal | 92.11 ± 0.60 | 89.57 ± 0.64 | 92.13 ± 0.64 | 87.93 ± 0.69 | 88.81 ± 0.47 | 90.38 ± 0.71 | 92.29 ± 0.70 | **92.44 ± 0.37** |
| Autos | 72.88 ± 2.76 | 68.39 ± 0.85 | 70.00 ± 2.91 | 61.41 ± 2.33 | 69.02 ± 1.51 | 70.05 ± 3.87 | 66.68 ± 0.98 | **75.76 ± 1.87** |
| Bands | 74.24 ± 2.04 | 63.54 ± 0.63 | **73.47 ± 1.64** | 59.94 ± 0.37 | 69.77 ± 2.20 | 69.38 ± 1.68 | 70.34 ± 2.19 | 72.39 ± 1.70 |
| Credit approval | 81.39 ± 0.53 | 84.86 ± 0.91 | 68.26 ± 1.15 | 71.55 ± 1.48 | 81.94 ± 0.83 | 82.32 ± 0.85 | 80.54 ± 0.96 | **85.33 ± 0.78** |
| German | 70.04 ± 1.40 | 67.98 ± 0.94 | 69.45 ± 0.75 | 64.57 ± 0.87 | 70.38 ± 1.27 | 70.16 ± 0.97 | – | **70.72 ± 1.05** |
| Heart disease | 76.56 ± 1.08 | 80.41 ± 1.51 | 71.52 ± 0.90 | 77.33 ± 1.46 | 74.30 ± 2.67 | 75.63 ± 1.55 | 76.15 ± 1.73 | **80.67 ± 0.83** |
| Sick | 98.85 ± 0.09 | 92.73 ± 0.18 | 98.78 ± 0.17 | 92.47 ± 0.17 | 98.84 ± 0.14 | 98.87 ± 0.16 | 98.82 ± 0.09 | **98.93 ± 0.08** |
| Dermatology | 93.96 ± 0.55 | 94.86 ± 0.40 | 94.84 ± 0.37 | 94.45 ± 0.34 | 79.21 ± 1.28 | 84.67 ± 0.95 | 92.32 ± 0.67 | **96.15 ± 0.37** |
| Ionosphere | 87.86 ± 0.97 | 85.81 ± 1.18 | 87.26 ± 1.20 | 85.01 ± 1.30 | 88.46 ± 1.41 | 88.97 ± 1.12 | 86.52 ± 1.51 | **89.46 ± 1.27** |
| Movement_libras | 66.25 ± 1.27 | 51.00 ± 1.15 | 58.03 ± 1.92 | 53.36 ± 1.75 | 63.78 ± 1.53 | 58.81 ± 2.12 | 40.67 ± 1.28 | **69.39 ± 1.69** |
| Page_blocks | 96.36 ± 0.11 | 96.06 ± 0.11 | 96.33 ± 0.06 | 96.33 ± 0.15 | – | – | **96.59 ± 0.16** | 96.48 ± 0.19 |
| Parkinsons | 85.85 ± 1.49 | 85.69 ± 1.07 | 85.59 ± 2.21 | 83.90 ± 2.68 | 85.79 ± 2.20 | **88.46 ± 1.71** | 85.23 ± 0.96 | 88.21 ± 1.79 |
| Sonar | 71.11 ± 2.83 | 76.11 ± 2.44 | 66.87 ± 2.29 | 72.64 ± 3.13 | 72.55 ± 2.05 | 61.54 ± 2.17 | 55.48 ± 2.88 | **76.68 ± 2.85** |
| WDBC | 92.64 ± 0.55 | 92.78 ± 0.71 | 91.02 ± 0.67 | 92.36 ± 0.76 | 92.67 ± 0.69 | 93.27 ± 0.66 | **93.69 ± 0.65** | 93.59 ± 0.74 |
| Winequality_red | 59.79 ± 0.76 | 58.73 ± 1.27 | 60.41 ± 0.83 | 60.46 ± 0.91 | – | – | 60.10 ± 1.23 | **61.08 ± 0.98** |
| Winequality_white | 58.03 ± 0.41 | 57.81 ± 0.59 | 57.50 ± 0.46 | 57.87 ± 0.86 | – | – | 57.93 ± 0.58 | **58.29 ± 0.38** |
| Yeast | 52.45 ± 1.14 | 52.91 ± 0.72 | 53.01 ± 1.09 | 52.76 ± 1.02 | 52.63 ± 0.74 | 52.68 ± 0.91 | – | **53.36 ± 1.14** |
| Chess | 99.34 ± 0.11 | 98.87 ± 0.10 | 96.76 ± 0.17 | 97.22 ± 0.12 | 98.26 ± 0.10 | 98.28 ± 0.12 | 99.10 ± 0.10 | 99.09 ± 0.10 |
| Lymphography | 74.93 ± 1.84 | **78.51 ± 2.11** | 74.53 ± 2.61 | 74.80 ± 2.67 | 76.49 ± 1.68 | 76.55 ± 2.47 | 77.70 ± 1.74 | 77.91 ± 2.94 |
| Mushroom | 100.00 ± 0.00 | **100.00 ± 0.00** | 99.90 ± 0.03 | 98.92 ± 0.01 | **100.00 ± 0.00** | **100.00 ± 0.00** | **100.00 ± 0.00** | **100.00 ± 0.00** |
| Avergae | 78.95 | 78.35 | 77.47 | 76.26 | – | – | – | **80.30** |

**Table 8**
The comparison of classification accuracy of the algorithm NB on reduced data (%)

| Data set | Raw data | LS | FSFS | SPEC | USQR | UEBR | UFRFS | FRUAR |
|---|---|---|---|---|---|---|---|---|
| Abalone | 24.98 ± 0.16 | 25.41 ± 0.22 | 25.37 ± 0.17 | 24.83 ± 0.19 | – | – | – | **25.43 ± 0.14** |
| Anneal | 76.09 ± 0.13 | 76.05 ± 0.11 | 76.07 ± 0.30 | 76.13 ± 0.07 | 75.74 ± 0.48 | 74.77 ± 1.62 | 76.08 ± 0.04 | **76.15 ± 0.41** |
| Autos | 60.59 ± 1.96 | 58.10 ± 0.81 | 62.05 ± 1.52 | 56.10 ± 0.92 | 61.51 ± 0.90 | 60.73 ± 1.84 | 57.85 ± 1.13 | **68.29 ± 1.85** |
| Bands | 70.45 ± 0.19 | 64.41 ± 0.00 | 68.95 ± 0.06 | 60.41 ± 0.89 | 69.08 ± 0.55 | 68.00 ± 0.77 | 66.27 ± 0.97 | **70.70 ± 0.69** |
| Credit approval | 66.72 ± 0.52 | 82.65 ± 0.21 | 67.33 ± 0.52 | 64.11 ± 0.47 | 82.57 ± 0.47 | 82.65 ± 0.42 | 67.74 ± 0.61 | **84.88 ± 0.31** |
| German | 70.83 ± 0.28 | 69.97 ± 0.18 | 71.64 ± 0.36 | 70.65 ± 0.26 | 71.39 ± 0.19 | 71.89 ± 0.48 | – | **72.33 ± 0.42** |
| Heart disease | 79.93 ± 0.52 | 82.11 ± 0.39 | 78.19 ± 0.41 | 79.41 ± 0.72 | 76.85 ± 0.73 | 77.11 ± 1.03 | 79.78 ± 0.72 | **82.30 ± 0.60** |
| Sick | 93.72 ± 0.04 | **93.88 ± 0.00** | 93.73 ± 0.02 | 93.72 ± 0.04 | 93.74 ± 0.02 | 93.68 ± 0.04 | 93.73 ± 0.03 | 93.74 ± 0.01 |
| Dermatology | 81.31 ± 0.55 | 76.75 ± 0.71 | 72.98 ± 1.02 | 78.47 ± 0.48 | 74.75 ± 0.71 | 74.64 ± 1.34 | 71.64 ± 0.93 | **82.60 ± 0.60** |
| Ionosphere | 90.57 ± 0.43 | 77.92 ± 0.77 | 89.83 ± 0.50 | 80.17 ± 0.49 | 90.91 ± 0.37 | 90.17 ± 0.24 | 85.30 ± 0.69 | **93.11 ± 0.26** |
| Movement_libras | 67.56 ± 1.15 | 37.89 ± 1.00 | 50.11 ± 0.89 | 43.42 ± 1.16 | 66.47 ± 0.85 | 58.44 ± 0.71 | 29.75 ± 1.25 | **67.60 ± 0.66** |
| Page_blocks | 94.13 ± 0.09 | 93.09 ± 0.07 | 94.58 ± 0.05 | 93.48 ± 0.09 | – | – | 95.23 ± 0.08 | **95.32 ± 0.07** |
| Parkinsons | 74.87 ± 0.80 | 68.77 ± 0.78 | 84.00 ± 0.68 | 76.26 ± 0.77 | 80.51 ± 0.73 | 76.05 ± 0.69 | 82.41 ± 0.77 | 83.49 ± 1.13 |
| Sonar | 75.62 ± 1.18 | 75.00 ± 1.11 | 74.76 ± 1.25 | 73.99 ± 0.92 | 72.16 ± 1.17 | 66.06 ± 0.99 | 57.40 ± 0.91 | **80.87 ± 1.63** |
| WDBC | 93.70 ± 0.31 | **94.20 ± 0.22** | 89.72 ± 0.26 | 93.12 ± 0.28 | 93.20 ± 0.25 | 93.99 ± 0.22 | 93.33 ± 0.26 | 93.86 ± 0.35 |
| Winequality_red | 58.24 ± 0.32 | 56.34 ± 0.39 | 57.33 ± 0.44 | 59.00 ± 0.37 | – | – | 57.29 ± 0.32 | **59.50 ± 0.40** |
| Winequality_white | 48.43 ± 0.13 | 48.01 ± 0.14 | 49.28 ± 0.17 | 48.52 ± 0.15 | – | – | 49.94 ± 0.20 | **50.17 ± 0.18** |
| Yeast | 48.12 ± 0.48 | 47.90 ± 0.38 | 48.34 ± 0.35 | 47.96 ± 0.43 | 49.69 ± 0.55 | 49.51 ± 0.49 | – | **49.82 ± 0.49** |
| Chess | 57.25 ± 0.38 | 70.83 ± 0.17 | 70.36 ± 0.18 | 55.99 ± 0.27 | 66.90 ± 0.15 | 66.90 ± 0.16 | **76.19 ± 0.31** | 75.05 ± 0.22 |
| Lymphography | 80.74 ± 2.57 | 78.51 ± 0.89 | **80.88 ± 1.86** | 76.01 ± 1.97 | 79.86 ± 1.77 | 79.26 ± 2.02 | 80.20 ± 2.16 | 79.66 ± 2.28 |
| Mushroom | 93.60 ± 0.04 | 88.72 ± 0.05 | **96.34 ± 0.05** | 90.17 ± 0.05 | 95.32 ± 0.02 | 95.32 ± 0.01 | 94.19 ± 0.03 | 95.30 ± 0.02 |
| Average | 72.85 | 72.37 | 73.58 | 70.93 | – | – | – | **76.47** |

In addition, we can also find that in most cases, the larger the number of original features, the higher the reduction efficiency. The more decision classes there are, the lower the classification accuracy is usually. The size of the number of objects has no certain effect on performance.

What's more, both the replacement and discretization of data values may cause changes in the data structure, which leads to the loss of information in turn. On data sets with numerical attributes, algorithms USQR and UEBR show relatively poor classification accuracy. This is because the numerical attributes are discretized before the experiment. In addition, for data sets containing nominal attributes, replacing the nominal attributes with different integer values will also affect the classification effect of algorithms LS, FSFS, SPEC, and UFRFS. However, for the algorithm FRUAR, neither replacement nor discretization need to be done, so that the data can retain more real information. Therefore, the algorithm FRUAR has relatively better classification results.

In summary, the algorithm FRUAR achieves optimal classification accuracy in more than half cases. In addition, from the perspective of average classification accuracy, the FRUAR algorithm achieves better values in all four classification algo-

**Table 9**
The comparison of classification accuracy of the algorithm kNN on reduced data (%).

| Data set | Raw data | LS | FSFS | SPEC | USQR | UEBR | UFRFS | FRUAR |
|---|---|---|---|---|---|---|---|---|
| Abalone | 19.95 ± 0.27 | 20.60 ± 0.18 | **20.92 ± 0.27** | 20.05 ± 0.16 | – | – | – | 20.73 ± 0.25 |
| Anneal | 90.79 ± 0.37 | 90.70 ± 0.32 | **91.72 ± 0.42** | 87.44 ± 0.39 | 88.51 ± 0.50 | 89.76 ± 0.54 | 90.98 ± 0.22 | 90.14 ± 0.41 |
| Autos | 73.27 ± 1.50 | 65.61 ± 1.13 | 72.73 ± 1.04 | 58.29 ± 1.40 | 72.68 ± 1.40 | 73.51 ± 1.24 | 72.10 ± 1.12 | **80.15 ± 1.49** |
| Bands | 76.16 ± 0.84 | 63.48 ± 0.47 | **78.38 ± 0.66** | 43.58 ± 0.47 | 70.79 ± 0.77 | 67.29 ± 0.98 | 63.15 ± 0.73 | 73.99 ± 0.86 |
| Credit approval | 80.51 ± 0.53 | 79.65 ± 0.56 | 69.20 ± 0.79 | 66.78 ± 0.86 | 81.03 ± 0.66 | 80.49 ± 0.31 | 80.36 ± 0.48 | **81.54 ± 0.62** |
| German | 69.22 ± 0.35 | 66.35 ± 0.43 | 68.86 ± 0.47 | 64.06 ± 0.56 | 68.39 ± 0.67 | 70.02 ± 0.82 | – | **70.79 ± 0.57** |
| Heart disease | 75.70 ± 1.11 | 74.89 ± 0.78 | 73.89 ± 1.77 | 74.56 ± 0.97 | 71.33 ± 1.11 | 71.52 ± 0.96 | 75.15 ± 0.62 | **78.19 ± 0.66** |
| Sick | 96.32 ± 0.14 | **96.39 ± 0.11** | 95.99 ± 0.12 | 96.04 ± 0.12 | 96.14 ± 0.09 | 96.24 ± 0.06 | 96.14 ± 0.13 | 96.35 ± 0.13 |
| Dermatology | 95.66 ± 0.27 | 95.36 ± 0.29 | 93.11 ± 0.53 | 95.77 ± 0.27 | 81.80 ± 0.41 | 85.38 ± 0.50 | 93.74 ± 0.33 | **96.12 ± 0.31** |
| Ionosphere | 86.89 ± 0.63 | 87.98 ± 0.58 | **90.46 ± 0.49** | 87.86 ± 0.31 | 88.66 ± 0.60 | 89.40 ± 0.32 | 87.81 ± 0.50 | 90.17 ± 0.36 |
| Movement_libras | 85.83 ± 0.63 | 53.03 ± 1.19 | 58.00 ± 0.91 | 51.61 ± 0.84 | 84.81 ± 0.94 | 83.86 ± 0.69 | 57.08 ± 1.14 | **87.19 ± 0.69** |
| Page_blocks | 95.99 ± 0.11 | 96.02 ± 0.09 | 95.97 ± 0.10 | 96.36 ± 0.08 | – | – | 96.01 ± 0.08 | **96.03 ± 0.10** |
| Parkinsons | 95.95 ± 0.78 | 96.36 ± 0.89 | 96.21 ± 0.97 | 90.62 ± 0.97 | 93.69 ± 0.73 | 94.77 ± 0.53 | 92.87 ± 0.85 | **96.62 ± 0.50** |
| Sonar | 86.39 ± 1.06 | 84.42 ± 0.94 | 84.90 ± 0.79 | 85.38 ± 0.97 | 79.71 ± 0.71 | 72.74 ± 1.06 | 55.34 ± 1.31 | **88.27 ± 0.82** |
| WDBC | 95.36 ± 0.24 | 94.62 ± 0.24 | 94.76 ± 0.31 | 94.31 ± 0.42 | 93.50 ± 0.31 | 95.33 ± 0.33 | 94.04 ± 0.31 | **96.31 ± 0.40** |
| Winequality_red | 64.50 ± 0.61 | 64.72 ± 0.56 | 64.80 ± 0.42 | 65.57 ± 0.55 | – | – | 64.36 ± 0.62 | **65.68 ± 0.39** |
| Winequality_white | 65.69 ± 0.33 | 64.99 ± 0.39 | 65.03 ± 0.30 | 64.45 ± 0.36 | – | – | **65.90 ± 0.24** | 65.75 ± 0.27 |
| Yeast | 52.82 ± 0.44 | 52.48 ± 0.56 | 52.45 ± 0.35 | **52.86 ± 0.46** | 52.67 ± 0.50 | 52.47 ± 0.41 | – | 52.64 ± 0.58 |
| Chess | 84.58 ± 0.18 | 83.95 ± 0.23 | 83.36 ± 0.13 | **88.66 ± 0.28** | 84.12 ± 0.19 | 84.17 ± 0.25 | 84.33 ± 0.28 | 84.02 ± 0.28 |
| Lymphography | 75.00 ± 1.35 | 69.80 ± 2.11 | 76.35 ± 1.56 | 69.39 ± 1.01 | **80.07 ± 1.16** | 79.59 ± 1.77 | 75.27 ± 0.65 | 79.12 ± 1.48 |
| Mushroom | 100.00 ± 0.00 | **100.00 ± 0.00** | 99.84 ± 0.03 | 98.58 ± 0.02 | 99.98 ± 0.02 | 99.99 ± 0.01 | **100.00 ± 0.00** | **100.00 ± 0.01** |
| Average | 79.63 | 78.14 | 79.05 | 75.85 | – | – | – | **80.76** |

rithms. The algorithm FRUAR can obtain a relatively small subset of attributes and improve or maintain classification accuracy. Therefore, the algorithm FRUAR is suitable for the reduction of classification with mixed attributes.

### 4.3. Clustering experiments

Second, we analyze the results of the clustering experiments. The comparison results of clustering experiments are shown in Tables 10 and 11. Among them, Table 10 presents the optimal attribute subset and number of the algorithm FRUAR regarding different clustering algorithms. Table 11 gives the comparison of the optimal clustering accuracy based on algorithm k-Means. The bold parts show the best clustering accuracy of different attribute reduction algorithms. "–" means that no reduction result is obtained.

Through Table 10, we can see that for all data sets, the algorithm FRUAR also can remove some redundant or irrelevant attributes. From the results in Table 11, it can be seen that the algorithm FRUAR has 13 cases to achieve the best clustering accuracy. However, for algorithms LS, FSFS, SPEC, USQR, UEBR, and UFRFS, only 1, 2, 2, 1, 0, and 2 cases achieve the best accu-

**Table 10**
The optimal attribute subsets and number of the algorithm k-Means

| Data set | Raw attribute | k-Means(Number) | $\lambda$ |
|---|---|---|---|
| Abalone | 8 | 2,1,4,6,8,7,3(7) | 1 |
| Anneal | 38 | 3,33,12,32,8,34,35,5,36,9,37,1,17(13) | 1 |
| Autos | 25 | 2,24,6,22,11,5,21,19,23,1(10) | 2.1 |
| Bands | 39 | 2,3,14,26,38(5) | 0.8 |
| Credit approval | 15 | 4,6,14,2,3,10,12,7,1,9,8(11) | 2.2 |
| German | 20 | 4,7,12,3,1,6,9,17,19,14,10,15(12) | 0.4 |
| Heart disease | 13 | 5,10,12,3,4,8,1,11,7,13,2,9(12) | 1.3 |
| Sick | 29 | 29,26,24,1,19,2,23,3,20,11,10,17,6,14,16,22,8,13,9,4,5,12,18,21,7(25) | 1 |
| Dermatology | 34 | 21,33,11,28,4,9,15,16,5,26,7,18,32,24,14,8,6,10,22,3,17,29,30,13,20,34(26) | 0.3 |
| Ionosphere | 33 | 26,25,14,13,27,7,2,29,3,33,32,9,5,8,17,15,24,19,6,10,11,4(22) | 1.1 |
| Movement_libras | 90 | 21,53,6,58,79,74,32,90,1,46,35,63,89,20,64,82,36,41,71,15,27,54,26,49,40(25) | 0.5 |
| Page_blocks | 10 | 9,5,6,2,4,1,10,8(8) | 0.5 |
| Parkinsons | 22 | 14,18,19,21,1,20,17,3,2(9) | 2.2 |
| Sonar | 60 | 58,48,13,33,37,5,60,52(8) | 2.4 |
| WDBC | 31 | 24,19,31,6,3,30,13,10,16,26,28,20,2,12,23,11,17,1,29,9,27,25(22) | 0.9 |
| Winequality_white | 11 | 5,3,2,9,10,8,6,1,7,11(10) | 2.3 |
| Yeast | 8 | 6,5,7,2,8,3,4(7) | 0.6 |
| Chess | 36 | 35,11,9,7,34,10,5,6,13,24,33,15,36,18,26,12,20,21,4,23,22,30,16,31,1,27,17,2,8,3,19,25,28(33) | – |
| Lymphography | 18 | 18,14,12,1,15,2,16,8,11,3(10) | – |
| Mushroom | 22 | 14,9,5,22,3,1,2,21,15,20,13,12,17,7,6(15) | – |
| Average | 27.3 | 14.3 | – |

**Table 11**
The comparison of clustering accuracy of the algorithm k-Means on reduced data (%).

| Data set | Raw data | LS | FSFS | SPEC | USQR | UEBR | UFRFS | FRUAR |
|---|---|---|---|---|---|---|---|---|
| Abalone | 15.10 ± 0.68 | 15.42 ± 0.51 | 15.35 ± 0.71 | 15.13 ± 0.67 | – | – | – | **15.76 ± 0.76** |
| Anneal | 40.29 ± 3.60 | 38.42 ± 3.56 | **44.49 ± 5.99** | 39.26 ± 3.54 | 39.92 ± 4.67 | 36.80 ± 5.21 | 38.50 ± 4.78 | 42.19 ± 4.12 |
| Autos | 40.98 ± 3.69 | 40.05 ± 3.08 | 40.05 ± 2.66 | 40.93 ± 2.27 | 41.85 ± 3.68 | 43.90 ± 5.15 | 40.88 ± 3.40 | **44.20 ± 2.58** |
| Bands | 56.87 ± 2.69 | 58.61 ± 4.08 | 58.04 ± 3.26 | 56.93 ± 1.37 | 58.76 ± 0.00 | 58.59 ± 0.58 | 54.24 ± 1.19 | **62.82 ± 0.16** |
| Credit approval | 74.59 ± 11.23 | 70.57 ± 13.70 | 78.35 ± 5.64 | 52.74 ± 0.99 | 78.23 ± 5.60 | 71.48 ± 13.41 | 78.75 ± 5.98 | **79.16 ± 10.08** |
| German | 52.20 ± 0.00 | 52.20 ± 0.00 | 57.55 ± 7.13 | **63.55 ± 1.62** | 53.77 ± 2.31 | 53.47 ± 2.28 | – | 56.07 ± 6.23 |
| Heart disease | 76.78 ± 6.67 | 72.70 ± 9.61 | 75.44 ± 6.46 | 75.33 ± 8.47 | 63.19 ± 7.49 | 65.19 ± 6.64 | 76.07 ± 7.59 | **80.37 ± 0.00** |
| Sick | 80.66 ± 8.74 | 80.09 ± 8.26 | 79.17 ± 10.39 | 79.35 ± 6.92 | 76.25 ± 6.44 | 74.82 ± 6.46 | 73.83 ± 7.12 | **82.92 ± 8.16** |
| Dermatology | 71.58 ± 11.79 | 68.42 ± 8.16 | 70.52 ± 12.20 | 71.78 ± 13.85 | 60.66 ± 7.27 | 65.85 ± 6.64 | 77.16 ± 7.16 | **77.27 ± 9.78** |
| Movement_libras | 71.23 ± 0.00 | 70.94 ± 0.00 | **72.93 ± 1.29** | 68.09 ± 0.00 | 68.66 ± 0.00 | 70.37 ± 0.00 | 72.11 ± 3.65 | 71.25 ± 0.00 |
| Ionosphere | 60.93 ± 8.78 | 64.77 ± 8.13 | 61.21 ± 7.59 | 58.79 ± 8.23 | 58.93 ± 3.57 | 58.13 ± 5.74 | 57.66 ± 6.78 | **64.92 ± 9.74** |
| Page_blocks | 51.46 ± 9.29 | 49.25 ± 6.39 | 49.40 ± 6.02 | **63.62 ± 6.08** | – | – | 49.22 ± 6.09 | 51.74 ± 5.13 |
| Parkinsons | 63.08 ± 0.00 | 59.85 ± 10.39 | 76.82 ± 0.32 | 68.72 ± 0.00 | 69.85 ± 2.12 | 64.92 ± 3.33 | 76.36 ± 0.85 | **77.44 ± 0.00** |
| Sonar | 54.90 ± 1.15 | 50.96 ± 0.00 | 56.88 ± 0.23 | 57.93 ± 0.25 | **68.75 ± 0.00** | 57.40 ± 2.16 | 56.25 ± 0.00 | 65.87 ± 0.00 |
| WDBC | 92.79 ± 0.00 | **93.50 ± 0.00** | 88.05 ± 0.00 | 91.74 ± 0.00 | 92.62 ± 0.00 | 92.65 ± 0.27 | 88.51 ± 9.11 | 92.97 ± 0.00 |
| Winequality_red | 30.58 ± 1.81 | 31.20 ± 2.61 | 29.70 ± 2.55 | 31.78 ± 2.17 | – | – | 30.59 ± 2.99 | **32.88 ± 1.25** |
| Winequality_white | 26.48 ± 1.07 | 27.80 ± 2.08 | 25.41 ± 1.90 | 28.01 ± 1.31 | – | – | 28.71 ± 2.39 | **28.96 ± 2.53** |
| Yeast | 38.81 ± 4.14 | 39.99 ± 1.35 | 39.95 ± 3.31 | 39.49 ± 2.42 | 39.43 ± 2.29 | 38.14 ± 1.08 | – | **40.28 ± 3.02** |
| Chess | 53.32 ± 3.15 | 52.59 ± 3.08 | 52.88 ± 3.90 | 53.72 ± 3.60 | 52.25 ± 2.43 | 51.10 ± 0.00 | **54.62 ± 4.55** | 54.61 ± 3.33 |
| Lymphography | 48.31 ± 4.61 | 47.64 ± 4.90 | 46.62 ± 6.96 | 47.91 ± 5.89 | 49.39 ± 2.28 | 48.38 ± 2.67 | **49.70 ± 2.74** | 49.68 ± 2.39 |
| Mushroom | 72.51 ± 14.10 | 74.99 ± 15.04 | 73.28 ± 14.06 | 70.89 ± 8.32 | 71.15 ± 8.99 | 74.32 ± 7.67 | 74.42 ± 14.80 | **75.88 ± 6.97** |
| Average | 55.10 | 53.90 | 55.88 | 54.99 | – | – | – | **58.46** |

racy, respectively. For the algorithm FRUAR, there are 20 cases higher than algorithms LS, USQR, and UEBR, 18 cases higher than algorithms FSFS and UFRFS, and 19 cases higher than the algorithm SPEC.

In summary, the algorithm FRUAR achieves better clustering accuracy. The optimal clustering accuracy is achieved in most cases on the algorithm k-Means. From the perspective of average clustering accuracy, the algorithm FRUAR achieves the maximum value on the algorithm k-Means. The algorithm FRUAR can provide a relatively small subset of attributes and improve or maintain clustering accuracy. Therefore, the algorithm FRUAR is suitable for the reduction of mixed attributes when used in clustering.

### *4.4. Outlier experiments*

Finally, we analyze the results of the outlier detection experiments. The experimental results are shown in Tables 12–15. Among them, Table 12 presents the optimal attribute subset and number of the algorithm FRUAR regarding different outlier detection algorithms. Tables 13–15 give the comparison of outlier detection results of algorithms DIS, kNNO, and LOF on reduced data, respectively.

From Table 12, we can see that the algorithm FRUAR can remove some candidate attributes for all outlier detection algorithms. In general, different outlier detection algorithms may require different subsets of attributes to obtain the best performance. However, the optimal attribute subsets of the three outlier detection algorithms are the same on Hepa and Pima data sets.

**Table 12**
The subset and number of optimal reduced attribute for outlier detection algorithms..

| Data set | Raw data | DIS | $\lambda$ | kNNO | $\lambda$ | LOF | $\lambda$ |
|---|---|---|---|---|---|---|---|
| Cre | 15 | 6,4,9,12,7,1,10,13,2,3,14,8 (12) | 0.6 | 6,4,9,12,7,1,10,13,2,3,14,8 (12) | 0.6 | 4,6,2,14,3,12,1,10,8,9 (10) | 2.5 |
| Heart | 13 | 3,13,11,7,2,6,9,12 (8) | 0.1 | 3,8,4,5,2,11,1,13,7,12,10 (11) | 1.5 | 3,4,8,5,2,11,1,13,7,12,9 (11) | 1.3 |
| Hepa | 19 | 6,4,5,10,8,12,9,1,3,11,7,14,13 (13) | 0.1 | 6,4,5,10,8,12,9,1,3,11,7,14,13 (13) | 0.1 | 6,4,5,10,8,12,9,1,3,11,7,14,13 (13) | 0.1 |
| Diab | 8 | 4,5,1,8,7,3,2 (7) | 0.1 | 2,6,7,5,1,8,3 (7) | 0.3 | 3,6,2,1,7,4,5 (7) | 0.9 |
| Iono | 34 | 28,11,27,4,8,18,3,34,5,10, 22,29,32,33,12,24,26,7 (18) | 2.0 | 28,23,9,4,18,24,34,33,5,10,25,3 (12) | 2.9 | 28,11,27,4,8,18,3,34,5,10, 29,22,32,7,30,26,24,33 (18) | 2.1 |
| Pima | 9 | 9,5,1,7,8,3,2,6 (8) | 0.1 | 9,5,1,7,8,3,2,6 (8) | 0.1 | 9,5,1,7,8,3,2,6 (8) | 0.1 |
| Sonar | 60 | 47,9,33,23,28,21,56,7,50,60,54,55 (12) | 1.5 | 49,60,30,37,56,1,55 (7) | 2.5 | 47,9,33,23,28,21,56,7,50,60,54,55 (12) | 1.5 |
| Wbc | 9 | 5,3,7,4,1,8,9,6 (8) | 0.7 | 9,5,2,6,4,8,1,7 (8) | 0.3 | 5,3,7,4,1,8,9,6 (8) | 0.7 |
| Lymp | 18 | 18,14,12,1,15,2,16,8,11,3(10) | – | 18,14,12,1,15,2,16,8,11,3 (10) | – | 18,14,12,1,15,2,16,8,11,3 (10) | – |
| Average | 19.1 | 10.1 | – | 9.1 | – | 10.2 | – |

**Table 13**
The comparison of outlier detection results on DIS (%).

| Data set | P | | FPR | | FNR | |
|---|---|---|---|---|---|---|
| | Raw data | Reduced data | Raw data | Reduced data | Raw data | Reduced data |
| Cre | **73.81** | 69.05 | **2.87** | 3.39 | **26.19** | 30.95 |
| Heart | 75.00 | **81.25** | 2.67 | **2.00** | 25.00 | **18.75** |
| Hepa | **88.89** | **88.89** | **1.18** | **1.18** | **11.11** | **11.11** |
| Diab | 42.31 | **46.15** | 3.00 | **2.80** | 57.69 | **53.85** |
| Iono | 91.67 | **100.00** | 0.89 | **0.00** | 8.33 | **0.00** |
| Pima | 49.09 | **50.91** | 5.60 | **5.40** | 50.91 | **49.09** |
| Sonar | 80.00 | **90.00** | 2.06 | **1.03** | 20.00 | **10.00** |
| Wbc | 87.18 | **92.31** | 1.13 | **0.68** | 12.82 | **7.69** |
| Lymp | 66.67 | **66.67** | **1.41** | **1.41** | **33.33** | **33.33** |
| Average | 72.73 | **76.14** | 2.31 | **1.99** | 27.27 | **23.86** |

**Table 14**
The comparison of outlier detection results on kNNO (%)

| Data set | P | | FPR | | FNR | |
|---|---|---|---|---|---|---|
| | Raw data | Reduced data | Raw data | Reduced data | Raw data | Reduced data |
| Cre | 64.29 | **66.67** | 3.92 | **3.66** | 35.71 | **33.33** |
| Heart | **81.25** | **81.25** | **2.00** | **2.00** | **18.75** | **18.75** |
| Hepa | **88.89** | 66.67 | **1.18** | 3.53 | **11.11** | 33.33 |
| Diab | 46.15 | **50.00** | 2.80 | **2.60** | 53.85 | **50.00** |
| Iono | 100.00 | **100.00** | **0.00** | **0.00** | **0.00** | **0.00** |
| Pima | **58.18** | **58.18** | **4.60** | **4.60** | **41.82** | **41.82** |
| Sonar | **90.00** | **90.00** | **1.03** | **1.03** | **10.00** | **10.00** |
| Wbc | 87.18 | **89.74** | 1.13 | **0.90** | 12.82 | **10.26** |
| Lymp | 66.67 | **100.00** | 1.41 | **0.00** | 33.33 | **0.00** |
| Average | 75.85 | **78.06** | **2.01** | 2.04 | 24.15 | **21.94** |

**Table 15**
The comparison of outlier detection results on LOF (%).

| Data set | P | | FPR | | FNR | |
|---|---|---|---|---|---|---|
| | Raw data | Reduced data | Raw data | Reduced data | Raw data | Reduced data |
| Cre | **30.95** | **30.95** | **7.57** | **7.57** | **69.05** | **69.05** |
| Heart | 75.00 | **81.25** | 2.67 | **2.00** | 25.00 | **18.75** |
| Hepa | **88.89** | **88.89** | **1.18** | **1.18** | **11.11** | **11.11** |
| Diab | 53.85 | **61.54** | 2.40 | **2.00** | 46.15 | **38.46** |
| Iono | 91.67 | **100.00** | 0.89 | **0.00** | 8.33 | **0.00** |
| Pima | 50.91 | **52.73** | 5.40 | **5.20** | 49.09 | **47.27** |
| Sonar | 80.00 | **90.00** | 2.06 | **1.03** | 20.00 | **10.00** |
| Wbc | 92.31 | **92.31** | **0.68** | **0.68** | **7.69** | **7.69** |
| Lymp | 100.00 | 66.67 | **0.00** | 1.41 | **0.00** | 33.33 |
| Average | 73.73 | **73.81** | 2.54 | **2.34** | 26.27 | **26.19** |

On the nine data sets, it can be seen from Tables 13–15 that the algorithm FRUAR has 8, 8, and 8 cases to maintain or improve the detection precision of raw data on algorithms DIS, kNNO, and LOF, respectively. Among them, there are 6, 5, and 7 cases to improve the detection precision of raw data on algorithms DIS, kNNO, and LOF, respectively. In addition, in most cases, the algorithm FRUAR can maintain or reduce the FPR and FNR of raw data. What's more, for the average value, the algorithm FRUAR can also get better results in most cases.

The related ROC curves are shown in Fig. 1. From Fig. 1, we can see that the algorithm DIS is the closest one to the upper left corner of the first quadrant on reduced data sets Iono, Sonar, and Wbc, and the area under the curve is the largest one. It shows that the performance of the algorithm DIS on reduced data is significantly better than the performance on the original data. For other reduced data, the algorithm DIS is close to the upper left corner of the first quadrant, and the area under the curve is relatively large. It indicates that the performance of the algorithm DIS on the reduced data is comparable to or slightly weaker than the performance on the original data. Similarly, it can be obtained that the performance of algorithms kNNO and LOF are significantly better than the performance on the original data on reduced data on Cre, Heart, Heart, Diab,

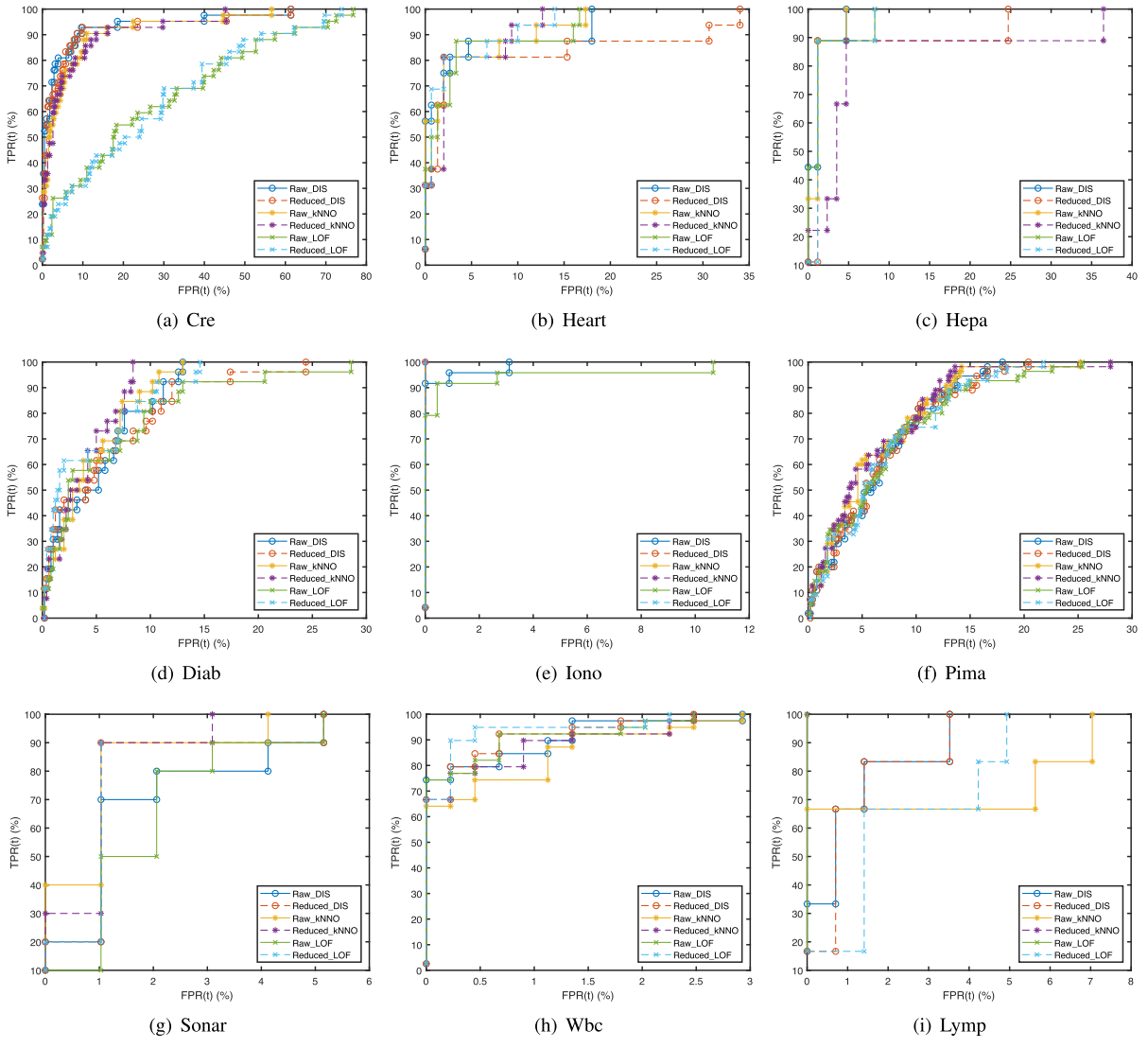



**Fig. 1.** The ROC curves of outlier detection algorithms on nine data sets.

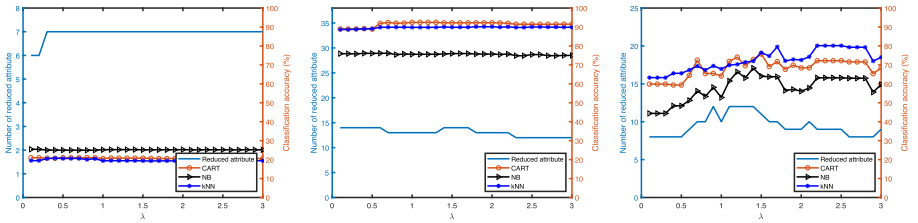
Pima, Lymph and Heart, Diab, Iono, Pima, Sonar, Lymph, respectively. Besides, their performance on other reduced data is comparable to or slightly weaker than that on the original data. Therefore, the algorithm DIS can better identify outliers in the most reduced data sets.

In conclusion, the algorithm FRUAR can reduce irrelevant or redundant attributes and improve or maintain the performance of outlier detection.

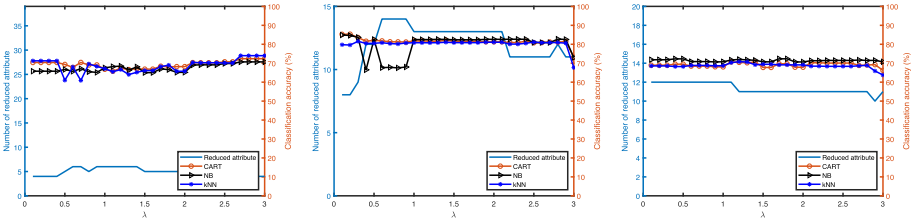
### *4.5. Parameter experiments*

The threshold $\lambda$ plays an important role in the algorithm FRUAR. It can be used as a parameter to control the fuzzy rough granules of data analysis. We can get different attribute subsets at each level of granularity. Based on this, a subset of attributes describing the recognition problems of different classification algorithms is obtained. The number of reduced attribute and classification accuracy curves with different $\lambda$ values are shown in Fig. 2. The following analyses are carried out from two aspects: the number of reduced attributes and classification accuracy.

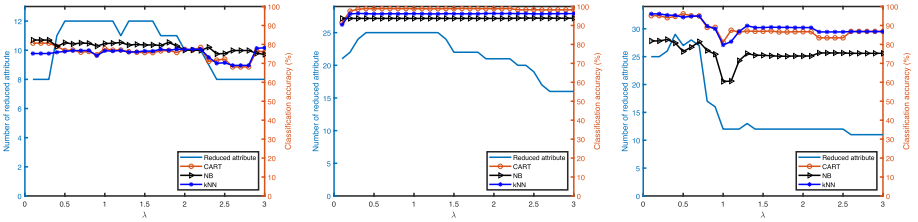
Through Fig. 2, we can see that on some data sets, as $\lambda$ increases, the number of reduced attributes increases to the maximum value, then decreases gradually, and it finally balances, such as Abalone, Parkinsons, and Sonar data sets. However, for Autos, Credit approval, and Heart disease data sets, the number of reduced attributes varies greatly with increasing $\lambda$. In



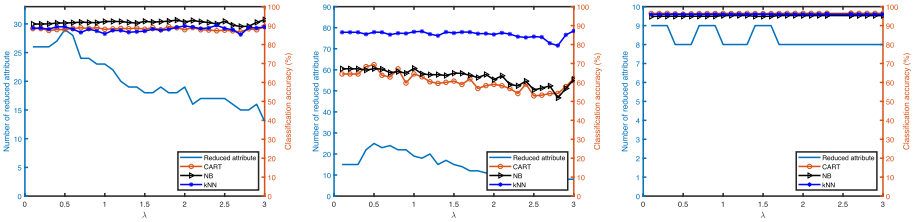

(a) Abalone (b) Anneal (c) Autos



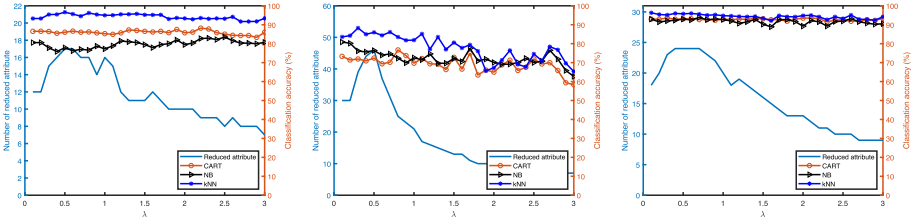

(d) Bands (e) Credit approval (f) German



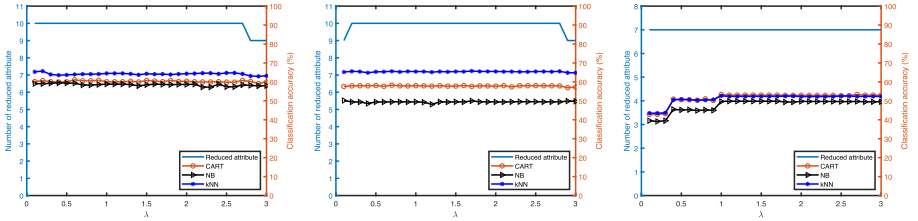

(g) Heart disease (h) Sick (i) Dermatology




(j) Ionosphere (k) Movement_libras (l) Page_blocks




(m) Parkinsons (n) Sonar (o) WDBC




(p) Winequality_red (q) Winequality_white (r) Yeast

**Fig. 2.** The number of reduced attribute and classification accuracy with different $\lambda$ values.

addition, on the Ionosphere data set, with $\lambda$ increases, the number of reduced attributes first reaches the maximum value, then decreases and starts to fluctuate, and finally does not stabilize.

As for classification accuracy, we can see from Fig. 2 that, for most data sets, the optimal value can be obtained under the value of multiple parameters $\lambda$. For each data set, we can choose the appropriate $\lambda$ according to Fig. 2. Therefore, the algorithm FRUAR is feasible for the attribute reduction of the classification algorithm.

For the parameters of classification and outlier detection, some similar conclusions can be made. the algorithm FRUAR is also feasible for the mixed attribute reduction of the clustering and outlier detection.

### 4.6. Hypothesis testing

Further, Friedman's test [15] and Nemenyi's post hoc test [9] are employed to evaluate the statistical significance of the algorithm. Before using Friedman's test, the classification accuracy of each algorithm on all datasets is sorted from low to high, and the sequence number is assigned $(1, 2, \cdots)$. Among them, if the classification accuracy of the two algorithms is the same, the ordinal values are equally divided. Then, Friedman's test is used to determine whether these algorithms have the same performance. Suppose we compare $M$ algorithms on $N$ data sets, and let $r_i$ represent the average ordinal value of the $i$th algorithm, then the Friedman's test is calculated as follows.

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(M-1)-\tau_{\chi^2}} \text{and} \tau_{\chi^2} = \frac{12N}{M(M+1)}\left(\sum_{i=1}^{M} r_i^2 - \frac{M(M+1)^2}{4}\right). \tag{24}$$

$\tau_F$ obeys the $F$ distribution with $(M-1)$ and $(M-1)(N-1)$ degrees of freedom. If the null hypothesis of "all algorithms have the same performance" is rejected, it means that the performance of the algorithms is significantly different. At this time, a post hoc test is needed to further distinguish these attribute reduction algorithms on learning algorithms. Nemenyi's post hoc test is commonly used. In Nemenyi's test, the critical difference (CD) of the average ordinal value is calculated by the following formula.

$$CD_\alpha = q_\alpha \sqrt{\frac{M(M+1)}{6N}}, \tag{25}$$

where $q_\alpha$ is the critical value of Tukey's distribution, which can be found in [9].

Given a significance level $\alpha$, the corresponding critical difference $CD_\alpha$ can be obtained. By calculating the difference of average ordinal value between any two reduction algorithms on a learning algorithm, if the difference between them is greater than $CD_\alpha$, the test result is that the performance of the two algorithms is significantly different on the learning algorithm.

However, the above methods cannot visually represent the significant differences between the two algorithms. For this reason, the above-mentioned test comparison is presented through Nemenyi's test figure. In Nemenyi's test figure, the average ordinal values of seven attribute reduction algorithms are plotted at the corresponding positions on the number line. If a group of algorithms is connected by horizontal line segments, then it means that there is no significant difference between this group of algorithms.

Specifically, in order to facilitate comparison, all rows without values in Tables 7–9 and 11 are deleted. Finally, we can get $M = 7$ and $N = 15$, the $\tau_F$ distribution has 6 and 84 degrees of freedom. According to Friedman's test, $\tau_F$ of different learning algorithms and critical value (significance level $\alpha$ is 0.1) are shown in Table 16. According to Table 16, when $\alpha = 0.1$, each values of $\tau_F$ on four learning algorithms is greater than the critical value 1.8455. Therefore, the null hypothesis that "all algorithms have the same classification performance" is rejected on four learning algorithms. It shows that the performance of all reduction algorithms is significantly different on each learning algorithm. At this time, a post hoc test is needed to further distinguish these attribute reduction algorithms on four learning algorithms.

For significance level $\alpha = 0.1$, the corresponding critical difference $CD_{0.1} = 2.1243$ can be obtained. Finally, Nemenyi's test figures on four learning algorithms are shown in Fig. 3. From Fig. 3, we can see that the algorithm FRUAR is statistically significantly different from most other algorithms. For example, it can be seen from Fig. 3(b) and 3(d) that the algorithm FRUAR is not connected by any horizontal line segment, which shows that the algorithm FRUAR and other algorithms are statistically significantly different on NB and k-Means classifiers. Fig. 3(a) demonstrates that the algorithm FRUAR is statistically better than algorithms LS, FSFS, SPEC, UFRFS, and USQR on the CART classifier, respectively. However, there is no consistent

**Table 16**
$\tau_F$ of different learning algorithms.

| Learning algorithms | $\tau_F$ | Critical value ($\alpha = 0.1$) |
|---|---|---|
| CART | 9.2227 | 1.8455 |
| NB | 5.1573 | |
| kNN | 4.5860 | |
| k-Means | 5.9390 | |

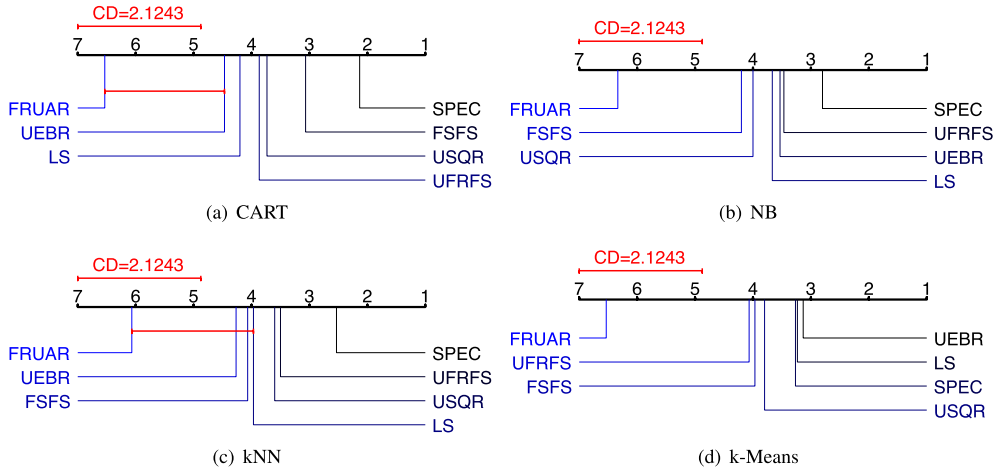



**Fig. 3.** Nemenyi's test figures on four learning algorithms.

evidence to indicate the statistical differences from the algorithm UEBR on the CART classifier. Besides, the average order value of four learning algorithms in Fig. 3 is also relatively large, which also shows the effectiveness of the algorithm FRUAR.

## 5. Conclusion

Aiming at the problem of unsupervised mixed attribute reduction, this paper proposes a generalized unsupervised mixed attribute reduction based on fuzzy rough sets. This method does not need to discretize the numerical attribute, which can reduce data preprocessing time and avoid information loss. Aiming at this method, a specific algorithm FRUAR is designed. Based on thirty UCI data sets, the algorithm FRUAR is compared with existing algorithms. Experimental results show that the algorithm can select fewer attributes to maintain or improve the ability of classification, clustering, and outlier detection for mixed attribute data. Furthermore, the hypothesis statistical test shows that the proposed algorithm is statistically significantly different from most other existing algorithms. In future work, unsupervised attribute reduction based on three-way decision models can be further studied.

## CRediT authorship contribution statement

**Zhong Yuan:** Conceptualization, Methodology, Software, Investigation, Writing - original draft. **Hongmei Chen:** Software, Resources, Project administration, Funding acquisition, Writing - review & editing. **Tianrui Li:** Supervision, Resources, Funding acquisition. **Zeng Yu:** Formal analysis, Data curation, Software. **Binbin Sang:** Formal analysis, Data curation, Software. **Chuan Luo:** Formal analysis, Data curation, Software.

## 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

This work was supported by the National Natural Science Foundation of China (61976182, 62076171, and 61572406), 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] C.C. Aggarwal, Outlier Analysis, Springer, 2016.
[2] R.B. Bhatt, M. Gopal, Frct: fuzzy-rough classification trees, Pattern Anal. Appl. 11 (1) (2008) 73–88.
[3] M.M. Breunig, H.P. Kriegel, R.T. Ng, J. Sander, Lof: identifying density-based local outliers, Acm Sigmod Record 29 (2) (2000) 93–104.
[4] G.O. Campos, A. Zimek, J. Sander, R.J. Campello, B. Micenková, E. Schubert, I. Assent, M.E. Houle, On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study, Data Min. Knowl. Disc. 30 (4) (2016) 891–927.
[5] D.G. Chen, L. Zhang, S.Y. Zhao, Q.H. Hu, P.F. Zhu, A novel algorithm for finding reducts with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 20 (2) (2012) 385–389.
[6] D.G. Chen, X.X. Zhang, X.Z. Wang, Y.J. Liu, Uncertainty learning of rough set-based prediction under a holistic framework, Inf. Sci. 463 (2018) 129–151.

[7] J.H. Dai, H. Hu, W.Z. Wu, Y.H. Qian, D.B. Huang, Maximal discernibility pairs based approach to attribute reduction in fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (4) (2018) 2174–2187.
[8] M. Dash, H. Liu, Consistency-based search in feature selection, Artif. Intell. 151 (1–2) (2003) 155–176.
[9] J. Demšar. Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research, 7(Jan):1–30, 2006..
[10] D. Dheeru, E. Taniskidou Karra (2017). http://archive.ics.uci.edu/ml.
[11] L.J. Dong, D.G. Chen, N. Wang, Z.H. Lu, Key energy-consumption feature selection of thermal power systems based on robust attribute reduction with rough sets, Inf. Sci. 532 (2020) 61–71.
[12] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. General Syst. 17 (2–3) (1990) 191–209.
[13] D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, in: Intelligent Decision Support, Springer, 1992, pp. 203–232.
[14] J.G. Dy, C.E. Brodley. Feature selection for unsupervised learning. J. Mach. Learning Res., 5(Aug):845–889, 2004..
[15] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[16] A. Ganivada, S.S. Ray, S.K. Pal, Fuzzy rough sets, and a granular neural network for unsupervised feature selection, Neural Networks 48 (2013) 91–108.
[17] X.F. He, D. Cai, P. Niyogi. Laplacian score for feature selection. In Advances in neural information processing systems, pages 507–514, 2006..
[18] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recogn. 40 (12) (2007) 3509–3521.
[19] Q.H. Hu, D.R. Yu, W. Pedrycz, D.G. Chen, Kernelized fuzzy rough sets and their applications, IEEE Trans. Knowl. Data Eng. 23 (11) (2010) 1649–1667.
[20] Q.H. Hu, D.R. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recogn. Lett. 27 (5) (2006) 414–423.
[21] R. Jensen, Q. Shen, Fuzzy–rough attribute reduction with application to web categorization, Fuzzy Sets Syst. 141 (3) (2004) 469–485.
[22] E.M. Knorr, R.T. Ng, V. Tucakov, Distance-based outliers: algorithms and applications, VLDB J. 8 (3) (2000) 237–253.
[23] 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.
[24] N. Mac ParthaláIn, R. Jensen, Unsupervised fuzzy-rough set-based dimensionality reduction, Inf. Sci. 229 (2013) 106–121.
[25] J.S. Mi, Y. Leung, H.Y. Zhao, T. Feng, Generalized fuzzy rough sets determined by a triangular norm, Inf. Sci. 178 (16) (2008) 3203–3213.
[26] P. Mitra, C. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 301–312.
[27] P. Ni, S.Y. Zhao, X.Z. Wang, H. Chen, C.P. Li, E.C. Tsang, Incremental feature selection based on fuzzy rough sets, Inf. Sci. 536 (2020) 185–204.
[28] H.C. Peng, F.H. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[29] S. Ramaswamy, R. Rastogi, K. Shim, Efficient algorithms for mining outliers from large data sets, ACM SIGMOD Record 29 (2) (2000) 427–438.
[30] K. Selvakumar, M. Karuppiah, L. SaiRamesh, S.H. Islam, M.M. Hassan, G. Fortino, K.-K.R. Choo, Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in wsns, Inf. Sci. 497 (2019) 77–90.
[31] B.Z. Sun, W.M. Ma, X. Xiao, Three-way group decision making based on multigranulation fuzzy decision-theoretic rough set over two universes, Int. J. Approximate Reasoning 81 (2017) 87–102.
[32] C. Velayutham, K. Thangavel, Unsupervised quick reduct algorithm using rough set theory, J. Electronic Sci. Technol. 9 (3) (2011) 193–201.
[33] 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), pages 156–161. IEEE, 2012..
[34] 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.
[35] C.Z. Wang, Y. Huang, M.W. Shao, X.D. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[36] C.Z. Wang, Y. Wang, M.W. Shao, Y.H. Qian, D.G. Chen, Fuzzy rough attribute reduction for categorical data, IEEE Trans. Fuzzy Syst. 28 (5) (2020) 818–830.
[37] H.J. Wang, Y.H. Zhang, J. Zhang, T.R. Li, L.X. Peng, A factor graph model for unsupervised feature selection, Inf. Sci. 480 (2019) 144–159.
[38] S.H. Wang, J.L. Tang, and H. Liu. Embedded unsupervised feature selection. In Twenty-ninth AAAI conference on artificial intelligence, pages 470–476, 2015..
[39] X.Z. Wang, E.C.C. Tsang, S.Y. Zhao, D.G. Chen, D.S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set technique, Inf. Sci. 177 (20) (2007) 4493–4514.
[40] D.S. Yeung, D.G. Chen, E.C.C. Tsang, J.W.T. Lee, X.Z. Wang, On the generalization of fuzzy rough sets, IEEE Trans. Fuzzy Syst. 13 (3) (2005) 343–361.
[41] D.R. Yu, Q.H. Hu, W. Bao, Combining rough set methodology and fuzzy clustering for knowledge discovery from quantitative data, Proc. CSEE 24 (6) (2004) 205–210.
[42] 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 and Systems, doi: 10.1016/j.fss.2020.10.017..
[43] 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. Applied Soft Computing, page 107353, 2021..
[44] Z. Yuan, X.Y. Zhang, S. Feng, Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures, Expert Syst. Appl. 112 (2018) 243–257.
[45] P.F. Zhang, T.R. Li, G.Q. Wang, C. Luo, H.M. Chen, J.B. Zhang, D.X. Wang, Z. Yu, Multi-source information fusion based on rough set theory: A review, Inform. Fusion 68 (2021) 85–117.
[46] 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 Recogn. 56 (2016) 1–15.
[47] X.Y. Zhang, H. Yao, Z.Y. Lv, D.Q. Miao, Class-specific information measures and attribute reducts for hierarchy and systematicness, Inf. Sci. 563 (2021) 196–225.
[48] 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.
[49] P.F. Zhu, Q. Xu, Q.H. Hu, C.Q. Zhang, Co-regularized unsupervised feature selection, Neurocomputing 275 (2018) 2855–2863.
[50] P.F. Zhu, W.C. Zhu, Q.H. Hu, C.Q. Zhang, W.M. Zuo, Subspace clustering guided unsupervised feature selection, Pattern Recogn. 66 (2017) 364–374.