



Attribute reduction in an incomplete categorical decision information system based on fuzzy rough sets

Jiali He¹ · Liangdong Qu²  · Zhihong Wang³ · Yiyang Chen⁴ · Damei Luo⁵ · Ching-Feng Wen^{6,7}

Published online: 18 January 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Categorical data is an important class of data in machine learning. Information system based on categorical data is called a categorical information system (CIS), a CIS with missing values is known as an incomplete categorical information system (ICIS) and an ICIS with decision attributes is said to be an incomplete categorical decision information system (ICDIS). Attribute selection is an important subject in rough set theory. This paper investigates attribute reduction in an ICDIS based on fuzzy rough sets. To depict the similarity for incomplete categorical data, fuzzy symmetry relations in an ICDIS are first introduced. Then, some attribute-evaluation functions, such as fuzzy positive regions, dependency function and attribute importance functions are given. Next, the fuzzy-rough iterative computation model for an ICDIS is presented, and an attribute reduction algorithm in an ICDIS based on fuzzy rough sets is given. Finally, experiments are carried out as so to evaluate the performance of the proposed algorithm, and Friedman test and Bonferroni–Dunn test in statistics are conducted. The experimental results indicate that the proposed algorithm is more effective than some existing algorithms.

Keywords Attribute reduction · ICDIS · Fuzzy rough set

1 Introduction

1.1 Research background

Rough set theory (RST), initiated by Pawlak (1982, 1991), is a significant method to deal with imprecision, fuzziness and uncertainty. Its biggest advantage is that it doesn't require any prior information other than the data set to process the problem (Cattaneo et al. 2016; Chen et al. 2017; Huang et al. 2017; Li et al. 2016; Lang et al. 2017; Tran et al. 2018; Tan et al. 2018; Yao and Zhang 2017). RST is based on classification

✉ Liangdong Qu
quliangdong100@126.com

✉ Ching-Feng Wen
chingfengwen100@126.com; cfwen@kmu.edu.tw

Extended author information available on the last page of the article

mechanism, which regards classification as equivalence relation in a specific space, and equivalence relation constitutes space partition. The main idea of this method is to use the knowledge in the known knowledge base to describe the imprecise or uncertain knowledge. It is widely known that it can effectively deal with the uncertainty of an information system (IS). In recent years, RST has attracted many researchers' attention, and its application is mostly related to an IS (Li et al. 2020, 2020, 2019; Xie et al. 2019; Yu et al. 2019; Zhang et al. 2017, 2018).

Fuzzy rough set theory is a mathematical theory proposed by Dubois and Prade (1990). Chen et al. (2005) systematically studied fuzzy rough set theory and its related applications. Moresi and Yankout (1998) gave axiomatic definition for fuzzy rough sets. Radzikowska and Kerre (2002) investigated a comparative study of fuzzy rough sets. Wang et al. (2016) proposed a fitting model for attribute reduction with fuzzy rough sets. Chen et al. (2012) a novel attribute reduction algorithm based on fuzzy rough sets. Hu et al. (2011) developed kernelized fuzzy rough sets and their applications.

1.2 The related works

Categorical data is one of the most common types of data in people's life, and it is an important class of data in machine learning. In order to overcome the deficiency of classical rough sets processing classification data, several improved models are proposed. Ziarko (1993) introduced a variable precision rough set model (VPRSM). Yao (2008) proposed probabilistic rough set approximations. Duntsch and Gediga (1998) studied uncertainty measures of VPRSM. Liang et al. (2014) presented a group incremental approach to attribute reduction applying VPRSM. Liang et al. (2013) came up with an accelerator for attribute reduction based on perspective of objects and attributes in VPRSM. Slezak (2002) brought up approximate entropy reducts of VPRSM.

Attribute reduction aims to remove redundant attributes in the calculation process, solve the complexity of high-dimensional data calculation and boost its accuracy. There is no doubt that attribute reduction is one of the focuses and hotspots of rough set theory from beginning to end. Up to now, there have been many outstanding results. Dai et al. (2013) presented attribute attribute based on conditional entropies for incomplete decision systems. Meng and Shi (2009) proposed a fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets. Yao and Zhang (2017) brought up class-specific attribute reduction in RST. Zhao and Qin (2014) put forward mixed attribute reduction in incomplete decision table. Teng et al. (2010) came up with attribute reduction algorithm based on conditional entropy under incomplete information system. Cornelis et al. (2010) obtained a generalized model of attribute reduction based on fuzzy tolerance relation within the context of fuzzy rough set theory; Wang et al. (2019) constructed fuzzy rough set-based attribute reduction using distance measures; Giang et al. (2020) gave an incremental method for attribute reduction of dynamic decision tables by using fuzzy partition distance and hybrid filter-wrapper, which obtains better classification ability than other methods; Chen et al. (2019) introduced a fuzzy kernel-attribute reduction method for heterogeneous data; Dai et al. (2018) applied maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets; Liu et al. (2020) defined the notions of accurate reduction and reduced invariant matrix, which are used to attribute reduction in an IS.

1.3 Motivation and inspiration

In the real world, there are mostly missing data sets. Wang et al. (2020) studied attribute reduction for categorical data and did not consider the case of incomplete categorical data. In order to solve practical problems more effectively, fuzzy symmetry relations in characterizing incomplete categorical data are generalized in this paper. First, fuzzy tolerance relations applicable to both complete and incomplete data are defined. Then a fuzzy rough iterative computing model with stronger generalization ability is established. Finally, this model are applied in some real data sets to test the performance effect of the model. Therefore, it is meaningful to study attribute reduction in an ICDIS based on fuzzy rough sets.

In this paper, a fuzzy rough iterative calculation model is derived theoretically. Then a reduction algorithm FR-IC is designed by using the model. The advantage of this algorithm is that an iterative formula is defined to reduce the attributes of incomplete information systems. By iterating the fuzzy relation matrix continuously, the maximum dependency of attributes is found, and the reduction set is obtained. With the increase of iteration times, the calculation formula of dependency is dynamically adjusted to ensure the convergence of the algorithm. The algorithm is fast and does not occupy too much memory. Through a large number of numerical experiments and comparison with other algorithms, it is found that the proposed algorithm is obviously better than some existing algorithms. The results are shown in the tables and images drawn in this paper.

1.4 Organization

This paper is organized as follows. In Sect. 2, some basic concepts of fuzzy relations and ICDISs are reviewed, and in order to depict the similarity for incomplete categorical data, fuzzy symmetry relations in an ICDIS are introduced. In Sect. 3, some attribute-evaluation functions including fuzzy positive regions, dependency functions and attribute importance functions are given. In Sect. 4, the fuzzy-rough iterative computation model for an ICDIS is presented. In Sect. 5, attribute reduction in an ICDIS based on fuzzy rough sets is studied and the corresponding algorithm is proposed. In Sect. 6, an attribute reduction algorithm in an ICDIS based on fuzzy rough sets is proposed and its complexity is analyzed. In Sect. 7, some experiments are carried out and the performance of the proposed algorithm is evaluated. In Sect. 7, this paper is concluded.

The framework of this paper is revealed in Fig. 1.

2 Preliminaries

In this section, some notions of fuzzy relations and ICDISs are reviewed, and fuzzy symmetry relations in an ICDIS are defined.

Throughout this paper, U and A signify two finite sets.

Put

$$U = \{u_1, u_2, \dots, u_n\}, A = \{a_1, a_2, \dots, a_m\}.$$

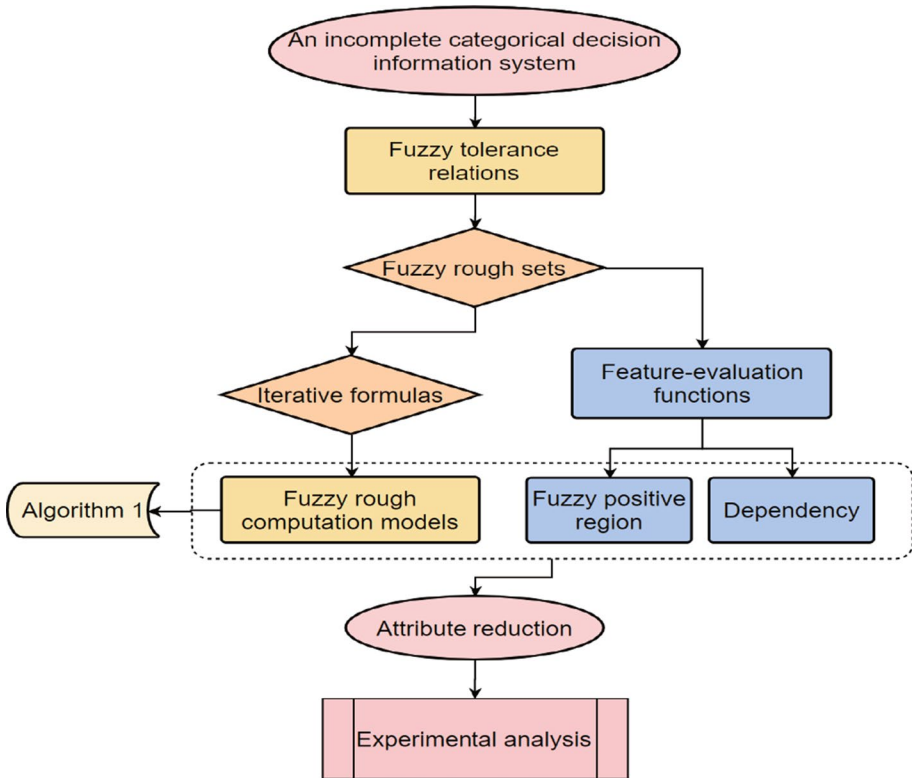


Fig. 1 Framework of this paper

2.1 Fuzzy relations

Fuzzy sets are extensions of ordinary sets (Zadeh 1965). A fuzzy set P in U is defined as a function assigning to each element u of U a value $P(u) \in I$ and $P(u)$ is called the membership degree of u to the fuzzy set P .

Throughout this paper, I expresses $[0, 1]$, I^U indicates the family of all fuzzy sets in U .

If R is a fuzzy set in $U \times U$, then R is called a fuzzy relation on U . In this paper, $I^{U \times U}$ denotes the set of all fuzzy relations on U .

Let $R \in I^{U \times U}$. Then R may be represented by

$$M(R) = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{pmatrix},$$

where $r_{ij} = R(u_i, u_j) \in I$ means the similarity between two objects u_i and u_j .

2.2 Incomplete categorical decision information systems

Definition 2.1 (Pawlak 1991) Suppose that U is a finite object set and A is a finite attribute set. Then the ordered pair (U, A) is referred to as an information system (IS), if for any $a \in C$, a is able to decide a information function $a : U \rightarrow V_a$, where $V_a = \{a(u) : u \in U\}$.

If $A = C \cup D$, where C is a set of conditional attributes and D is a set of decision attributes, then (U, A) is called a decision information system (DIS).

Let (U, A) be an IS. If there is $a \in A$ such that $* \in V_a$, here $*$ means a null or unknown value, then (U, A) is called an incomplete information system (IIS).

Suppose that (U, A) is an IIS. Given $P \subseteq A$. Then a binary relation T_P on U can be defined as

$$T_P \text{ or } \text{sim}(P) = \{(u, v) \in U \times U : \forall a \in P, a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *\}.$$

Clearly, T_P is a tolerance relation on U .

For each $u \in U$, denote

$$T_P(u) = \{v \in U : (u, v) \in T_P\}.$$

Then, $T_P(u)$ is called the tolerance class of x under the tolerance relation T_P .

For convenience, $T_{\{a\}}$ and $T_{\{a\}}(u)$ are denoted by T_a and $T_a(u)$, respectively.

Obviously,

$$T_P = \bigcap_{a \in P} T_a, \quad T_P(u) = \bigcap_{a \in P} T_a(u).$$

Let $(U, C \cup \{d\})$ be an DIS. If there exist $a \in C$ such that $* \in V_a$, but $* \notin V_d$, then $(U, C \cup \{d\})$ is called an incomplete decision information system (IDIS).

Let $(U, C \cup \{d\})$ be an IDIS. Then

$$T_d \text{ is an equivalence relation on } U.$$

Obviously,

$$\forall u \in U, T_d(u) = \{v \in U : d(u) = d(v)\}.$$

Definition 2.2 Let $(U, C \cup \{d\})$ be an IDIS. Then the pair (U, A) is called an incomplete categorical decision information system (ICDIS), if each attribute is categorical.

If $P \subseteq C$, then $(U, P \cup \{d\})$ is known as the subsystem of $(U, C \cup \{d\})$.

2.3 Fuzzy symmetry relations in an ICDIS

In T_P , " $\forall a \in P, a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *$ " is fed back to the object set of an IIS. Naturally, we may consider that " $\forall a \in P, a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *$ " is fed back to the attribute set of an IIS. For this purpose, inspired by the paper Wang et al. (2020), we introduce the following definition.

Definition 2.3 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$ and $\lambda \in [|P|, |C|]$. Then the fuzzy relation on U can be defined as

$$R_p^\lambda(u, v) = \frac{1}{\lambda} |\{a \in P : a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *\}|, \quad (2.1)$$

where $*$ is a missing value.

Moreover,

$$T_d = \{(u, v) \in U \times U : d(u) = d(v)\}.$$

Clearly,

R_p^λ is a fuzzy symmetry relation on U , and T_d is an equivalence relation on U . If $\lambda = |P|$, then R_p^λ is a fuzzy tolerance relation on U .

In this paper, denote

$$U/d = \{T_d(u) : u \in U\} = \{D_1, D_2, \dots, D_r\}.$$

For convenience, denote

$$P_{uv} = \{a \in P : a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *\}.$$

Then

$$R_p^\lambda(u, v) = \frac{1}{\lambda} |P_{uv}|.$$

Proposition 2.4 Let $(U, C \cup \{d\})$ be an ICDIS. If $P_1 \subseteq P_2 \subseteq C$, then for $\lambda \in [|P_2|, |C|]$, $R_{P_1}^\lambda \subseteq R_{P_2}^\lambda$.

Proof By Definition 2.3,

$$R_{P_1}^\lambda(u, v) = \frac{1}{\lambda} |(P_1)_{uv}|, \quad R_{P_2}^\lambda(u, v) = \frac{1}{\lambda} |(P_2)_{uv}|.$$

Since $P_1 \subseteq P_2$, we have $\forall u, v \in U, (P_1)_{uv} \subseteq (P_2)_{uv}$.

So $\forall u, v \in U, R_{P_1}^\lambda(u, v) \leq R_{P_2}^\lambda(u, v)$.

Thus, $R_{P_1}^\lambda \subseteq R_{P_2}^\lambda$. □

3 Attribute-evaluation functions

In this section, some attribute-evaluation functions, such as fuzzy positive regions, dependency functions and attribute importance functions are presented.

Definition 3.1 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$ and $\lambda \in [|P|, |C|]$. Suppose that R_p^λ is the fuzzy tolerance relation induced by the subsystem $\underline{U}(U, P \cup \{d\})$. Based on the fuzzy approximation space (U, R_p^λ) , a pair of operations $\underline{R}_p^\lambda, \overline{R}_p^\lambda: 2^U \rightarrow I^U$ are defined as follows:

$$\underline{R}_p^\lambda(X)(u) = \bigwedge_{v \notin X} [1 - R_p^\lambda(u, v)], \forall u \in U; \tag{3.1}$$

$$\overline{R}_p^\lambda(X)(u) = \bigvee_{v \in X} R_p^\lambda(u, v), \forall u \in U. \tag{3.2}$$

Then $\underline{R}_p^\lambda(X)$ and $\overline{R}_p^\lambda(X)$ are called the lower and upper fuzzy approximations of X , respectively.

Proposition 3.2 *Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P \subseteq C$ and $\lambda \in [|P|, |C|]$. Then the following properties hold.*

(1) $R_p^\lambda(\emptyset) = \underline{R}_p^\lambda(\emptyset) = \underline{0}$, $R_p^\lambda(U) = \overline{R}_p^\lambda(U) = \overline{1}$.

(2) $\underline{R}_p^\lambda(X) \subseteq X \subseteq \overline{R}_p^\lambda(X)$, where

$$X(u) = \begin{cases} 1, & u \in X; \\ 0, & u \notin X. \end{cases}$$

(3) $X \subseteq Y \Rightarrow \underline{R}_p^\lambda(X) \subseteq \underline{R}_p^\lambda(Y)$, $\overline{R}_p^\lambda(X) \subseteq \overline{R}_p^\lambda(Y)$.

(4) $\underline{R}_p^\lambda(X \cap Y) = \underline{R}_p^\lambda(X) \cap \underline{R}_p^\lambda(Y)$, $\overline{R}_p^\lambda(X \cap Y) = \overline{R}_p^\lambda(X) \cup \overline{R}_p^\lambda(Y)$.

(5) $\underline{R}_p^\lambda(U - X) = \overline{1} - \overline{R}_p^\lambda(X)$,

$\overline{R}_p^\lambda(U - X) = \underline{1} - \underline{R}_p^\lambda(X)$.

Proof We only prove (2) and (5).

(2) (i) By Definition 2.3, $\forall u, v \in U, 0 \leq R_p^\lambda(u, v) \leq 1$.

Then

$$0 \leq \bigwedge_{v \notin X} [1 - R_p^\lambda(u, v)] \leq 1.$$

It can be obtained that $\forall u \in U$,

$$0 \leq \underline{R}_p^\lambda(X)(u) \leq 1.$$

So $\forall u \in X$,

$$\underline{R}_p^\lambda(X)(u) \leq 1 = X(u).$$

Note that $\forall u \notin X, \underline{R}_p^\lambda(X)(u) = 1 - R_p^\lambda(u, u) = 0$. Then $\forall u \notin X$,

$$\underline{R}_p^\lambda(X)(u) = 0 \leq 0 = X(u).$$

This implies that $\forall u \in U$,

$$\underline{R}_p^\lambda(X)(u) \leq X(u).$$

Thus

$$\underline{R}_p^\lambda(X) \subseteq X.$$

(ii) Obviously,

$$\forall u \in U, 0 \leq \overline{R_p^\lambda}(X)(u) \leq 1.$$

Then $\forall u \notin X$,

$$X(u) = 0 \leq \overline{R_p^\lambda}(X)(u).$$

Note that $\forall u \in X$, $\overline{R_p^\lambda}(X)(u) = R(u, u) = 1$. Then $\forall u \in U$,

$$X(u) = 1 \leq 1 = \overline{R_p^\lambda}(X)(u).$$

Thus

$$X \subseteq \overline{R_p^\lambda}(X).$$

From the above,

$$\underline{R_p^\lambda}(X) \subseteq X \subseteq \overline{R_p^\lambda}(X).$$

(5) (i) By Definition 3.1, we have $\forall u \in U$,

$$\begin{aligned} \underline{R_p^\lambda}(U - X)(u) &= \bigwedge_{v \notin U - X} [1 - R_p^\lambda(u, v)] \\ &= 1 - \bigvee_{v \notin U - X} R_p^\lambda(u, v) \\ &= 1 - \bigvee_{v \in X} R_p^\lambda(u, v) \\ &= 1 - \overline{R_p^\lambda}(X)(u). \end{aligned}$$

Thus

$$\underline{R_p^\lambda}(U - X) = \bar{1} - \overline{R_p^\lambda}(X).$$

(ii) By Definition 3.1, we have $\forall u \in U$,

$$\begin{aligned} \overline{R_p^\lambda}(U - X)(u) &= \bigvee_{v \in U - X} R_p^\lambda(u, v) \\ &= 1 - [1 - \bigvee_{v \in U - X} R_p^\lambda(u, v)] \\ &= 1 - [1 - \bigvee_{v \notin X} R_p^\lambda(u, v)] \\ &= 1 - \bigwedge_{v \notin X} [1 - R_p^\lambda(u, v)] \\ &= 1 - \underline{R_p^\lambda}(X)(u). \end{aligned}$$

Thus

$$\overline{R_p^\lambda}(U - X) = \bar{1} - \underline{R_p^\lambda}(X).$$

□

Proposition 3.3 *Let $(U, C \cup \{d\})$ be an ICDIS. If $P_1 \subseteq P_2 \subseteq C$, then for $\lambda \in [|P_2|, |C|]$ and $X \in 2^U$,*

$$\underline{R_{P_1}^\lambda}(X) \subseteq \underline{R_{P_2}^\lambda}(X), \overline{R_{P_2}^\lambda}(X) \subseteq \overline{R_{P_1}^\lambda}(X).$$

Proof (1) Since $P_1 \subseteq P_2 \subseteq C$, by Proposition 2.4, we have $R_{P_1}^\lambda \subseteq R_{P_2}^\lambda$.

Then $\forall u \in U, v \notin X$,

$$R_{P_1}^\lambda(u, v) \leq R_{P_2}^\lambda(u, v).$$

This implies that $\forall u \in U$,

$$\bigwedge_{v \notin X} [1 - R_{P_2}^\lambda(u, v)] \leq \bigwedge_{v \notin X} [1 - R_{P_1}^\lambda(u, v)].$$

Thus $\forall u \in U$,

$$\underline{R_{P_2}^\lambda}(X)(u) \leq \underline{R_{P_1}^\lambda}(X)(u).$$

Therefore,

$$\underline{R_{P_2}^\lambda}(X) \subseteq \underline{R_{P_1}^\lambda}(X).$$

(2) Since $P_1 \subseteq P_2 \subseteq C$, by Proposition 2.4, we have $R_{P_1}^\lambda \subseteq R_{P_2}^\lambda$.

Then $\forall u \in U, v \in X$,

$$R_{P_1}^\lambda(u, v) \leq R_{P_2}^\lambda(u, v).$$

This implies that $\forall u \in U$,

$$\bigvee_{v \in X} R_{P_1}^\lambda(u, v) \leq \bigvee_{v \in X} R_{P_2}^\lambda(u, v).$$

Thus $\forall u \in U$,

$$\overline{R_{P_1}^\lambda}(X)(u) \leq \overline{R_{P_2}^\lambda}(X)(u).$$

Therefore, $\overline{R_{P_1}^\lambda}(X) \subseteq \overline{R_{P_2}^\lambda}(X)$.

□

Suppose $D \in U/d$. Then

$$\underline{R_p^\lambda}(D), \overline{R_p^\lambda}(D) \in I^U.$$

$\underline{R_p^\lambda}(D)(u)$ denotes the membership degree of u certainly being included in the equivalence class D . If $u \notin D$, the value of $\underline{R_p^\lambda}(D)(u)$ is the smallest. Otherwise, it is equal to the smallest value of dissimilar degrees between u and the samples not falling into class D .

$\overline{R}_p^\lambda(D)(u)$ represents the membership degree of sample u possibly belonging to equivalence class D . If $u \notin D$, the value of $\overline{R}_p^\lambda(D)(u)$ is the largest. If not, it is equal to the max-value of the fuzzy similarities between u and all the samples in class D .

Definition 3.4 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P \subseteq C$ and $\lambda \in [|P|, |C|]$. Then λ -fuzzy positive region of decision d relative to P can be defined as

$$POS_p^\lambda(d) = \bigcup_{D \in U/d} R_p^\lambda(X). \tag{3.3}$$

Proposition 3.5 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P_1 \subseteq P_2 \subseteq C$. Given $\lambda \in [|P_2|, |C|]$. Then $POS_{P_2}^\lambda(d) \subseteq POS_{P_1}^\lambda(d)$.

Proof It follows from Proposition 3.2. □

Definition 3.6 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P \subseteq C$ and $\lambda \in [|P|, |C|]$. Then λ -dependency function of d relative to P can be defined as

$$\Gamma_p^\lambda(d) = \frac{|POS_p^\lambda(d)|}{n}. \tag{3.4}$$

Proposition 3.7 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P_1 \subseteq P_2 \subseteq C$. Given $\lambda \in [|P_2|, |C|]$. Then $\Gamma_{P_2}^\lambda(d) \leq \Gamma_{P_1}^\lambda(d)$.

Proof It follows from Proposition 3.5. □

Definition 3.8 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P \subseteq C$ and $a \in C - P$. Given $\lambda \in [|P|, |C|]$. Then λ -importance of a relative to P about d can be defined as

$$sig^\lambda(a, P, d) = \Gamma_p^\lambda(d) - \Gamma_{P \cup \{a\}}^\lambda(d). \tag{3.5}$$

4 Fuzzy rough computation models

Fuzzy relations introduced above are crucial for defining a fuzzy rough computation model. Noticed that there is a constant parameter λ . For a dataset with a large number of attributes, the membership degrees of samples to a relation can get very small when a few of the attributes are included in rough computation. That is to say, the lower the number of the included attributes in rough computation, the smaller the discrimination of memberships. To overcome this problem, a fuzzy rough iterative computation model for incomplete categorical data is proposed in this section.

Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Suppose $|P| \leq \lambda_1 < \lambda_2 < \dots \leq |C|$. Denote

$$R_p^{(i)}(u, v) = R_p^{\lambda_i}(u, v); \tag{4.1}$$

$$\underline{R}_p^{(i)}(D)(u) = \bigwedge_{v \notin D} [1 - R_p^{(i)}(u, v)] \quad (D \in U/d, u \in U); \tag{4.2}$$

$$\overline{R}_P^{(i)}(D)(u) = \bigvee_{v \in D} R_P^{(i)}(u, v) \quad (D \in U/d, u \in U); \tag{4.3}$$

$$POS_P^{(i)}(d) = \bigcup_{D \in U/d} \overline{R}_P^{(i)}(D); \tag{4.4}$$

$$\Gamma_P^{(i)}(d) = \frac{|POS_P^{(i)}(d)|}{n}; \tag{4.5}$$

$$sig^{(i)}(a, P, d) = \Gamma_P^{(i)}(d) - \Gamma_{P \cup \{a\}}^{(i)}(d) \quad (P \subseteq C, a \in C - P). \tag{4.6}$$

Obviously,

$$\begin{aligned} POS_P^{(i)}(d) &= POS_P^{\lambda_i}(d); \\ \Gamma_P^{(i)}(d) &= \Gamma_P^{\lambda_i}(d); \\ sig^{(i)}(a, P, d) &= sig^{\lambda_i}(a, P, d) \quad (P \subseteq C, a \in C - P). \end{aligned}$$

Theorem 4.1 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then $\forall D \in U/d, \forall u \in U, \forall i$,

$$\overline{R}_P^{(i+1)}(D)(u) = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \overline{R}_P^{(i)}(D)(u). \tag{4.7}$$

$$\begin{aligned} \overline{R}_P^{(i)}(D)(u) &= \bigwedge_{v \notin D} (1 - \frac{1}{\lambda_i} |P_{uv}|) \\ &= 1 - \bigvee_{v \notin D} \frac{1}{\lambda_i} |P_{uv}| \\ &= 1 - \frac{1}{\lambda_i} \bigvee_{v \notin D} |P_{uv}|. \end{aligned}$$

Proof

Similarly,

$$\overline{R}_P^{(i+1)}(D)(u) = 1 - \frac{1}{\lambda_{i+1}} \bigvee_{v \notin D} |P_{uv}|.$$

Then

$$\begin{aligned} \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \overline{R}_P^{(i)}(D)(u) &= \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \left(1 - \frac{1}{\lambda_i} \bigvee_{v \notin D} |P_{uv}| \right) \\ &= 1 - \frac{1}{\lambda_{i+1}} \bigvee_{v \notin D} |P_{uv}|. \end{aligned}$$

Thus

$$\overline{R_P^{(i+1)}}(D)(u) = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \overline{R_P^{(i)}}(D)(u).$$

□

Corollary 4.2 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$ and $\lambda \in [|P|, |A|]$. Then $\forall D \in U/d, \forall u \in U, \forall i \in [|P|, |C| - 1]$,

$$\overline{R_P^{(i+1)}}(D)(u) = \frac{1}{i+1} + \frac{i}{i+1} \overline{R_P^{(i)}}(D)(u). \quad (4.8)$$

Theorem 4.3 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$ and $\lambda \in [|P|, |A|]$ and $D \in U/d$. Then $\forall i$,

$$\overline{R_P^{(i+1)}}(D) = \frac{\lambda_i}{\lambda_{i+1}} \overline{R_P^{(i)}}(D). \quad (4.9)$$

Proof Obviously. □

Corollary 4.4 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$ and $D \in U/d$. Then $\forall i \in [|P|, |C| - 1]$,

$$\overline{R_P^{(i+1)}}(D) = \frac{i}{i+1} \overline{R_P^{(i)}}(D). \quad (4.10)$$

Theorem 4.5 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then $\forall u \in U, \forall i$,

$$POS_P^{(i+1)}(d)(u) = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} POS_P^{(i)}(d)(u). \quad (4.11)$$

Proof By Theorem 4.1, we can obtain that $\forall u \in U, \forall i$,

$$\begin{aligned} POS_P^{(i+1)}(d)(u) &= \bigvee_{D \in U/d} \overline{R_P^{(i+1)}}(D)(u) \\ &= \bigvee_{D \in U/d} \left(\frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \overline{R_P^{(i)}}(D)(u) \right) \\ &= \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \bigvee_{D \in U/d} \frac{\lambda_i}{\lambda_{i+1}} \overline{R_P^{(i)}}(D)(u) \\ &= \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \bigvee_{D \in U/d} \overline{R_P^{(i)}}(D)(u). \end{aligned}$$

Thus

$$POS_P^{(i+1)}(d)(u) = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} POS_P^{(i)}(d)(u).$$

□

Corollary 4.6 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then $\forall u \in U, \forall i \in [|P|, |C| - 1]$,

$$POS_P^{(i+1)}(d)(u) = \frac{1}{i+1} + \frac{i}{i+1} POS_P^{(i)}(d)(u). \quad (4.12)$$

Theorem 4.7 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then $\forall i$,

$$\Gamma_P^{(i+1)}(d) = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \Gamma_P^{(i)}(d). \quad (4.13)$$

Proof By Theorem 4.5,

$$\begin{aligned} |POS_P^{(i+1)}(d)| &= \sum_{i=1}^n POS_P^{(i+1)}(d)(u_i) \\ &= \sum_{i=1}^n \left(\frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} POS_P^{(i)}(d)(u_i) \right) \\ &= \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} n + \frac{\lambda_i}{\lambda_{i+1}} \sum_{i=1}^n POS_P^{(i)}(d)(u_i) \\ &= \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} n + \frac{\lambda_i}{\lambda_{i+1}} |POS_P^{(i)}(u_i)|. \end{aligned}$$

Then

$$\frac{|POS_P^{(i+1)}(d)|}{n} = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \frac{|POS_P^{(i)}(d)|}{n}.$$

Thus

$$\Gamma_P^{(i+1)}(d) = \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \Gamma_P^{(i)}(d).$$

□

Corollary 4.8 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then $\forall i \in [|P|, |C| - 1]$,

$$\Gamma_P^{(i+1)}(d) = \frac{1}{i+1} + \frac{i}{i+1} \Gamma_P^{(i)}(d). \quad (4.14)$$

Theorem 4.9 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P \subseteq C, a \in C - P$. Then $\forall i$,

$$sig^{(i+1)}(a, P, d) = \frac{\lambda_i}{\lambda_{i+1}} sig^{(i)}(a, P, d). \quad (4.15)$$

Proof By Theorem 4.7,

$$\begin{aligned}
 \text{sig}^{(i+1)}(a, P, d) &= \Gamma_P^{(i+1)}(d) - \Gamma_{P \cup \{a\}}^{(i+1)}(d) \\
 &= \frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \Gamma_P^{(i)}(d) - \left(\frac{\lambda_{i+1} - \lambda_i}{\lambda_{i+1}} + \frac{\lambda_i}{\lambda_{i+1}} \Gamma_{P \cup \{a\}}^{(i)}(d) \right) \\
 &= \frac{\lambda_i}{\lambda_{i+1}} \Gamma_P^{(i)}(d) - \frac{\lambda_i}{\lambda_{i+1}} \Gamma_{P \cup \{a\}}^{(i)}(d) \\
 &= \frac{\lambda_i}{\lambda_{i+1}} (\Gamma_P^{(i)}(d) - \Gamma_{P \cup \{a\}}^{(i)}(d)).
 \end{aligned}$$

Thus

$$\text{sig}^{(i+1)}(a, P, d) = \frac{\lambda_i}{\lambda_{i+1}} \text{sig}^{(i)}(a, P, d).$$

□

Corollary 4.10 Let $(U, C \cup \{d\})$ be an ICDIS. Suppose $P \subseteq C$, $a \in C - P$. Then $\forall i \in [|P|, |C| - 1]$,

$$\text{sig}^{(i+1)}(a, P, d) = \frac{i}{i+1} \text{sig}^{(i)}(a, P, d). \quad (4.16)$$

5 Attribute reduction in an ICDIS

In this section, attribute reduction in an ICDIS based on fuzzy rough sets is studied and the corresponding algorithm is proposed.

Definition 5.1 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then P is called a coordination subset of C relative to d , if $POS_P^\lambda(d) = POS_C^\lambda(d)$ where $\lambda = |P|$.

Definition 5.2 Suppose that $(U, C \cup \{d\})$ is an ICDIS. Given $a \in P \subseteq C$. Then a is called independent in P , if $POS_P^\lambda(d) \neq POS_{P-\{a\}}^\lambda(d)$ where $\lambda = |P|$.

Definition 5.3 Assume that $(U, C \cup \{d\})$ is an ICDIS. Given $P \subseteq C$. Then P is called independent, if for any $a \in P$, a is independent in P .

Definition 5.4 Let $(U, C \cup \{d\})$ be an ICDIS. Given $P \subseteq C$. Then P is called a reduct of C to d , if P is both coordination and independent.

In this paper, the family of all coordination subsets (resp., all reducts) of C to d is denoted by $co(C)$ (resp., $red(C)$).

Obviously,

$$P \in red(C) \iff P \in co(C) \text{ and } \forall P' \subset P, P' \notin co(C).$$

Theorem 5.5 Suppose that $(U, C \cup \{d\})$ is an ICDIS. Given $P \subseteq C$. Then

$$P \in co(C) \iff \Gamma_P^\lambda(d) = \Gamma_C^\lambda(d), \text{ where } \lambda = |P|.$$

Based on the discussion above, the following attribute reduction algorithm in an ICDIS based on fuzzy rough sets is proposed.

Algorithm 1: Fuzzy rough set algorithm for attribute reduction in an ICDIS. (FR-IC)

Input: An ICDIS $(U, C \cup \{d\})$ and $D \in U/d$.
Output: one reduct P .

- 1 Denote $n = |U|$, $m = |C|$. For any $a_k \in C$ ($k = 1, 2, \dots, m$), let $S_{\{a_k\}}(D) = (0)_{nn}$, $sv = (0)_{nn}$, $temp_{\{a_k\}} = (0)_{nn}$, $P = \emptyset$, $R_{\{a_k\}}(D) = (0)_{1n}$.
- 2 **for** each $u_i \in U$ **do**
- 3 **for** each $u_j \in U$, and $i \neq j$ **do**
- 4 **if** $d(u_i) = d(u_j)$ **then**
- 5 $sv(i, j) = NaN$.
- 6 **end**
- 7 **end**
- 8 **end**
- 9 **for** each $a_k \in C$ **do**
- 10 **for** each $u_i \in D$ **do**
- 11 **for** each $u_j \in U - D$ **do**
- 12 **if** $a_k(u_i) = a_k(u_j)$ or $a_k(u_i) = *$ **then**
- 13 $r(i, j) = 1$.
- 14 **else**
- 15 $r(i, j) = 0$.
- 16 **end**
- 17 **end**
- 18 **end**
- 19 $S_{\{a_k\}}(D) = (r(i, j))_{nn}$.
- 20 **end**
- 21 **while** $start=1$ **do**
- 22 $\lambda = |P|$.
- 23 **for** each $a_k \in C - P$ **do**
- 24 $temp_{\{a_k\}} = S_{\{a_k\}}(D) + sv$.
- 25 **for** each $i = 1, \dots, n$ **do**
- 26 Compute $R_{\{a_k\}}(D)(i) = \min\{1 - \frac{1}{\lambda+1} temp_{\{a_k\}}(i)\}$, where $temp_{\{a_k\}}(i)$ is the i th row vector of $temp_{\{a_k\}}$.
- 27 **end**
- 28 $\Gamma_{P \cup \{a_k\}}^\lambda(d) = \frac{sum\{R_{\{a_k\}}(D)\}}{n}$.
- 29 **end**
- 30 Find a^* with maximum value $\Gamma_{P \cup \{a_k\}}^\lambda(d)$.
- 31 Compute $sig(a^*, P, d) = \frac{\lambda}{\lambda+1} (\Gamma_P^\lambda(d) - \Gamma_{P \cup \{a^*\}}^\lambda(d))$.
- 32 **if** $sig(a^*, P, d) > \delta$ **then**
- 33 $P = P \cup \{a^*\}$;
- 34 $sv = sv + S_{\{a^*\}}(D)$.
- 35 **else**
- 36 **break**.
- 37 **end**
- 38 **end**
- 39 **Return** P .

By using Algorithm 1, the time complexity of attribute reduction in an ICDIS is polynomial. First, Steps 2 to 8 take $O(n^2)$ time complexity and Steps 9 to 20 is $O(mn^2)$. The time complexity of Step 23 to 29 is $O(mn^2)$, and that of Step 30 is $O(m)$. The time complexity of Steps 32 to 37 is $O(n^2)$, so the overall complexity from Step 21 to Step 38

is $O(m(m+1)n^2/2 + m^2 + mn^2) = O(m^2n^2)$. Thus the time complexity of Algorithm 1 is $O(n^2 + mn^2 + m^2n^2) = O(m^2n^2)$. Furthermore, its space complexity is $O(mn^2)$.

6 Experimental analysis

In this section, the performance of the proposed attribute reduction algorithm and existing algorithms is evaluated. The frame work chart of experiments is displayed in Fig. 2.

6.1 Datasets

Eight data sets are used in the experimental analysis, which are selected from UCI (Frank and Asuncion 2010) Machine Learning Repository. These data sets containing missing values are described in Table 1.

6.2 Data preprocessing

Data preprocessing composed of three main steps: Label Transfer and Data Transfer, Missing value processing, Remove duplication, and Data normalization. In label transfer and data transfer, all the symbolic data are transferred to numeric values. It is important to process missing value in the training set to calculate the distance between individuals. For categorical data, we usually use frequency values to deal with missing values. It is important to remove duplicate records in the training set to avoid the classifiers to be biased to most frequent records and prevent it from learning infrequent records. Normalizing the data is an important step to eliminate the biased with the features of larger

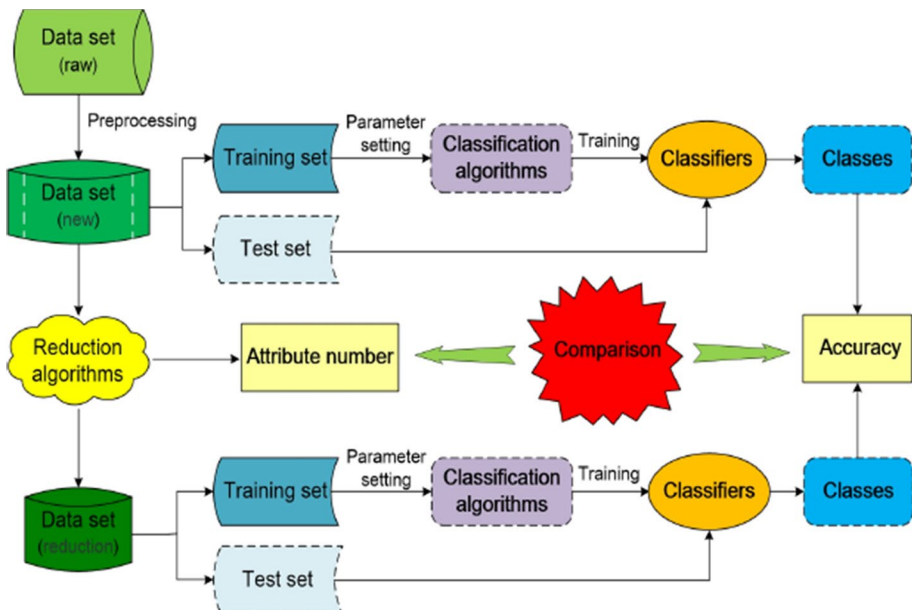


Fig. 2 The frame work chart of experiments

Table 1 The data sets excerpted from the UCI machine learning repository

No	Data sets	Abbreviation	Sample size	Attributes	Classes
1	Lung Cancer	LC	32	56	3
2	Large Soybean	LS	307	35	19
3	Dermatology	Der	358	34	6
4	Vote	Vot	435	16	2
5	Breast Cancer	BC	683	9	4
6	Mushroom	Mus	8124	22	2
7	Mammographic	Man	961	5	2
8	Audiology.standardiz	Aud	200	69	25

values from the data set. Data normalization is the process of transforming or scaling the data values of each feature into a proportional range. The used dataset was normalized into the range [0, 1] according to Eq.(6.1).

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}. \quad (6.1)$$

In this paper, because fuzzy relations are constructed directly and attribute reduction is performed based on fuzzy relations, there is no need to deal with missing values and normalize the data before attribute reduction.

6.3 Classifier training and testing

To evaluate these attribute reduction approaches, three learning mechanisms which can create classifiers are employed. They are frequently-used classifiers which the one is k-nearest nearest rule (KNN,K=3), the others are decision tree induction algorithms (ID3 and C4.5). We use confusion matrix to calculate the accuracy. As shown in Fig. 3.

$$accuracy = \frac{\sum_{i=1}^n c_{ii}}{\sum_{i=1}^n \sum_{j=1}^n c_{ij}}. \quad (6.2)$$

The proposed algorithm is compared with the other six algorithms. These are representative forward attribute reduction algorithms based on intuitionistic fuzzy rough (FMIFRS) Pankhuri et al. (2020), fuzzy similarity-based rough set (FSRS) Singh et al. (2020), dependency (SFFSNTCE) Zhao and Qin (2014), positive regions (PR) Meng and Shi (2009), conditional entropies (IE) Teng et al. (2010) and the heuristic search algorithm of SetCover approach (SetCover) Dai et al. (2013).

The parameter δ of FR-IC method is introduced to control the variable precision. The δ is increased from 0.05 to 0.25 with a step of 0.01. Because different values of δ produce different reduces, the experimental results with the highest classification accuracies are compared. In order to make the calculation more accurate, we did 10 experiments for each result. In each experiment, 20% was randomly selected as the test set and 80% as the training set, and then the classification accuracy calculated was averaged.

Confusion matrix		Predicted class			
		Class 1	Class 2	...	Class n
Actual class	Class 1	c_{11}	c_{12}	...	c_{1n}
	Class 2	c_{21}	c_{22}	...	c_{2n}
	⋮	⋮	⋮	\ddots	⋮
	Class n	c_{n1}	c_{n2}	...	c_{nn}

Fig. 3 Confusion matrix of classifier

The experiments are carried out on a personal computer. All attribute reduction algorithms are run in Matlab 2018b and hardware environment with an Intel(R)Core(TM) i7-9700CPU@3.0GHZ and 8GB RAM.

6.4 Classification results

Table 2 provides the average sizes of attribute reduction with these algorithms. Among these seven algorithms, only algorithm FMIFRFS deals with complete data. For this reason, the frequency mode method is used. The missing values in each column are replaced by the most frequent attribute values in this column. In algorithm FMIFRFS, the $\mu(x)$ function value is obtained by normalizing the data, and then the $v(x)$ function value is equal to $1-\mu(x)\text{-rand}/c$, where rand is a random number between (0,1) and c is a constant.

Table 2 Numbers of attribute reduction

Data sets	Raw data	FMIFRFS	FSRS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
Lung Cancer	56	31	5	3	7	4	4	4
Large Soybean	35	12	11	9	9	10	10	10
Dermatology	34	15	7	4	5	5	5	2
Vote	16	12	16	10	8	11	8	8
Breast Cancer	9	4	4	3	3	4	4	3
Mushroom	22	8	5	5	5	5	6	4
Mammographic	5	4	5	4	4	4	3	4
Audiology.standardiz	69	2	34	12	12	17	7	13
Average	30.75	11	10.88	6.25	6.63	7.5	5.88	6

It is easily seen that these attributes of data sets can be reduced effectively. From Table 2, it is easy to find out that the less average number of selected attributes are FR-IC and SETCOVER, which both are approximately equal to 6. The most number of selected attributes was carried out by FMIFRFS method, because its ending condition: The positive field of attribute reduction must be equal to 1, or the two values before and after are equal. Most data sets are difficult to achieve this condition for FMIFRFS, resulting in a large number of reduction. FSRS algorithm also requires that the positive field of the reduction attribute be equal to 1. Therefore, the positive field of all attributes in two datasets (Vote, Mammographic) cannot reach 1. So the reduction sets generated by FSRS algorithms are the original data for the two datasets.

The optimal attribute subsets with the highest classification accuracy of the seven algorithms are shown in Table 3. It is obvious to see that most of the attributes selected by these methods are not the same, because they may have more than one reduct. In Table 3, the “-” represents a gradual increase from one number to another.

Tables 4, 5 and 6 present the classification results with KNN ($k=3$), ID3 and C4.5. The underlined symbol denotes the highest classification accuracy among these attribute reduction algorithms. In Table 4, for Dermatology, FMIFRFS performs better than FR-IC in classifier KNN. The classification accuracy of FR-IC is a little worse than that of FSRS for Mushroom and Mammographic. FSRS algorithm does not complete attribute reduction in Vote and Mammographic, in other words, no attribute has been deleted, so the accuracy used is the original dataset. For the classifier KNN, the classification accuracy of three data is higher than that of the proposed algorithm in this paper.

As can be clearly seen from Table 5, for Dermatology, FR-IC is not good enough for FMIFRFS in classifier ID3. For Large Soybean and Mushroom, the classification accuracy of FSRS do better than other algorithms in ID3. From Table 6, for Mushroom, PR and SETCOVER execute better than FR-IC in classifier C4.5. The FMIFRFS algorithm performs better than other algorithms in Dermatology dataset. On the whole, classification accuracy based on the FR-IC method is higher than the other six methods in most case. We have done 168 numerical experiments, and 13 of them show that other algorithms are better than FR-IC algorithm. That is to say the excellent rate of FR-IC is 92.3% and it reaches a good effect. Therefore, our algorithm is superior to other six algorithms. In a word, the FR-IC method is more effective for attribute reduction in an ICDIS.

By discussing the influence of parameter δ on FR-IC algorithm, it is found that a reduct is related to δ , and the relationship between δ and classification accuracy is closely related. So it is easy to find the δ value of the optimal reduct. Figures 4, 5, 6, 7, 8, 9 and 10 show that the classification accuracy curve has no obvious change with the number of selected attributes decreasing. This shows that the proposed algorithm is effective. These figures show that the proposed algorithm has achieved good results. Figures 4, 5, 6, 7, 8, 9 and 10 show the relationship between the size of the reduct and the threshold δ , and the relationship between the classification accuracy of different classifiers and δ . The x-axis expresses the threshold δ , that value increases from 0.05 to 0.25 in 0.01 step. The left pink Y-axis indicates number of attributes of the reduct P , and the right blue Y-axis represents classification accuracy. The pink curve represents the relationship between δ and the size of the reduct P , and the function value of each point on the curve corresponds to the left pink y-axis. The other three curves with different colors reflect the classification accuracy of δ and three different classifiers, and the function value corresponds to the right blue y-axis. It is obvious that most of data sets can get higher classification accuracy in this experiment. Although classification accuracy curves of KNN and ID3 are gentle, the classification accuracy curve of C4.5 fluctuate obviously. The classification accuracy of C4.5 is lower

Table 3 The optimal attribute reduction among FMIFRFS, FRS, SFNSNCE, PR, IE, SETCOVER and FR-IC

Data sets	FMIFRFS	FRS	SFNSNCE	PR	IE	SETCOVER	FR-IC
LC	1, 5, 6, 8, 9, 12 17–33, 35, 36, 38 40, 43, 45, 47–56	40, 6, 12, 3, 54	6, 3, 12	34, 35, 42, 44 53, 55, 56	6, 3, 12, 15	13, 3, 12, 34	40, 6, 12, 3
LS	5, 11–13, 16, 20–22 24, 25, 27, 30	22, 27, 34, 4 16, 1, 32	1, 7, 6, 15, 22 17, 19, 4, 11	1, 7, 9, 10, 12, 14, 17, 22, 35	1, 7, 6, 15, 22, 17, 19, 4, 11, 12	1, 7, 6, 15, 22, 17, 19, 4, 12, 11	1, 12, 4, 15, 28, 22, 7, 6, 19, 11
Der	11, 13, 8, 22, 15, 5	22, 27, 34, 4 16, 1, 32	34, 4, 21, 3	19, 21, 28, 32, 34	34, 4, 21, 3, 32	34, 4, 21, 3, 32	22, 27
Vot	1, 3–16	1–16	4, 11, 3, 12, 9, 1, 14, 13, 10, 2	1, 2, 3, 4, 9, 10, 11, 15	4, 11, 3, 12, 9, 1, 14, 10, 13, 2, 15,	4, 11, 10, 9, 1, 12, 3, 14	1, 2, 3, 4 9, 10, 11, 15
BC	6, 2, 9, 4	2, 6, 1, 8	3, 1, 6	5, 6, 7	3, 1, 6, 2	3, 1, 6, 2	2, 1, 7
Mus	8, 12, 4, 7 13, 18, 19, 11	5, 20, 18, 13, 21	5, 9, 14, 20, 8	7, 5, 14, 6, 20	7, 5, 14, 6, 15	9, 5, 14, 6, 15, 18	5, 9, 14, 20
Mam	1, 5, 3, 2	1, 2, 3, 4, 5	3, 1, 2, 4	3, 1, 2, 5	3, 1, 2, 4	3, 1, 2	1, 2, 3, 5
Aud	11 22	1, 2, 6, 11–13, 15, 17, 21, 23, 28, 30, 31 33–36, 41, 44, 45, 47 50, 52, 55, 57, 59, 61, 63–69	64, 2, 60, 1 15, 5, 10, 66 38, 54, 6, 65	1, 2, 6, 57, 58, 59, 60, 62, 63, 64, 65, 66	64, 2, 60, 1, 15, 10, 66, 14, 54, 6 38, 59, 65 34, 11, 47	64, 2, 60, 1, 15, 5, 65	11, 43, 55 13, 63, 69 16, 2, 66 1, 15, 5, 64

Table 4 Comparison of classification accuracies of reduced data with KNN

Data sets	Raw data	FMIFRS	FSRS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
LC	51.34 ± 3.83	45.66 ± 1.48	52.37 ± 1.98	55.42 ± 7.78	47.50 ± 7.37	51.33 ± 8.02	47.90 ± 5.95	63.69 ± 1.56
LS	83.47 ± 1.91	69.19 ± 1.47	69.49 ± 2.84	65.37 ± 3.36	63.33 ± 1.85	64.28 ± 1.73	64.59 ± 1.98	70.40 ± 3.93
Der	87.25 ± 0.96	94.38 ± 2.13	68.27 ± 1.33	59.08 ± 4.16	58.72 ± 1.01	60.54 ± 1.11	61.10 ± 3.16	64.75 ± 0.96
Vot	90.47 ± 2.09	91.69 ± 0.70	90.47 ± 2.09	90.53 ± 1.69	91.39 ± 0.85	90.37 ± 1.15	91.59 ± 2.36	94.47 ± 0.71
BC	97.11 ± 0.70	94.34 ± 1.04	95.08 ± 1.40	95.36 ± 0.32	94.13 ± 1.21	95.92 ± 0.31	96.21 ± 0.43	96.37 ± 0.75
Mus	100.00 ± 0.00	98.69 ± 0.80	100.00 ± 0.00	99.51 ± 0.34	99.44 ± 0.26	98.91 ± 0.33	98.25 ± 0.40	99.52 ± 0.40
Mam	80.85 ± 0.25	73.24 ± 0.83	80.85 ± 0.25	76.71 ± 3.18	77.06 ± 1.59	77.32 ± 2.09	71.67 ± 0.47	77.56 ± 3.37
Aud	55.27 ± 5.02	27.28 ± 1.99	52.08 ± 2.43	52.46 ± 2.30	52.29 ± 1.25	53.70 ± 1.72	49.82 ± 4.78	54.25 ± 2.68
Average	80.72 ± 1.84	74.31 ± 1.31	73.36 ± 1.54	74.31 ± 3.27	72.98 ± 1.92	74.05 ± 2.06	75.93 ± 1.31	76.09 ± 1.99

Table 5 Comparison of classification accuracies of reduced data with ID3

Data sets	Raw data	FMFRFS	FORS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
LC	52.50 ± 2.39	54.15 ± 1.55	60.25 ± 1.55	43.12 ± 4.76	39.94 ± 9.26	50.17 ± 4.09	38.98 ± 3.44	64.42 ± 9.68
LS	80.93 ± 2.07	71.21 ± 2.33	81.87 ± 2.55	70.14 ± 2.49	63.10 ± 4.59	68.37 ± 3.44	68.45 ± 2.00	74.97 ± 2.87
Der	94.32 ± 0.49	95.42 ± 0.37	81.50 ± 0.10	69.48 ± 2.52	62.51 ± 6.69	70.63 ± 1.51	71.54 ± 3.24	65.19 ± 0.67
Vot	95.35 ± 0.78	93.65 ± 1.07	95.35 ± 0.78	95.45 ± 1.28	95.57 ± 1.24	95.52 ± 1.07	96.18 ± 0.44	96.38 ± 1.70
BC	93.75 ± 0.55	94.67 ± 0.37	94.94 ± 0.69	94.21 ± 0.51	94.42 ± 0.73	94.29 ± 0.48	94.46 ± 0.66	95.27 ± 1.43
Mus	100.00 ± 0.00	98.61 ± 0.13	99.95 ± 0.03	99.45 ± 0.18	99.53 ± 0.23	99.01 ± 0.27	99.06 ± 0.22	99.45 ± 0.31
Mam	81.87 ± 0.57	81.53 ± 1.18	81.87 ± 0.57	82.07 ± 0.58	82.27 ± 0.62	81.96 ± 0.94	81.82 ± 1.03	82.40 ± 0.12
Aud	69.22 ± 1.78	22.56 ± 1.31	67.39 ± 2.41	68.68 ± 2.75	64.13 ± 1.46	69.20 ± 2.16	59.04 ± 4.31	69.58 ± 3.61
Average	83.49 ± 1.08	76.48 ± 1.04	82.89 ± 1.04	77.83 ± 1.88	75.18 ± 3.10	82.71 ± 1.81	76.19 ± 1.92	82.58 ± 2.55

Table 6 Comparison of classification accuracies of reduced data with C4.5

Data sets	Raw data	FMIFRS	FIRS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
LC	44.52 ± 3.53	54.57 ± 1.78	54.89 ± 2.14	45.86 ± 2.56	42.65 ± 6.26	39.30 ± 4.81	36.90 ± 3.28	56.87 ± 4.71
LS	81.80 ± 1.83	48.73 ± 2.77	58.20 ± 1.30	73.28 ± 3.30	62.76 ± 0.64	72.71 ± 2.78	71.53 ± 2.48	76.08 ± 1.46
Der	30.25 ± 1.47	94.23 ± 0.72	33.71 ± 1.96	31.52 ± 1.70	30.01 ± 1.79	29.33 ± 3.13	30.21 ± 0.29	48.99 ± 1.06
Vot	95.40 ± 0.79	94.04 ± 1.73	95.40 ± 0.79	95.25 ± 0.55	95.59 ± 0.38	95.39 ± 0.92	95.18 ± 1.10	96.74 ± 1.79
BC	94.23 ± 1.06	91.21 ± 1.20	90.87 ± 0.47	91.36 ± 0.93	90.55 ± 1.43	90.47 ± 1.81	90.85 ± 1.11	92.39 ± 0.87
Mus	100.00 ± 0.00	98.83 ± 0.01	99.58 ± 0.08	98.74 ± 0.31	100.00 ± 0.00	98.91 ± 0.16	100.00 ± 0.00	98.85 ± 0.31
Mam	78.50 ± 1.44	76.20 ± 2.98	78.50 ± 1.44	79.22 ± 1.19	80.49 ± 0.74	79.21 ± 1.15	80.82 ± 2.05	81.08 ± 1.58
Aud	67.93 ± 2.72	23.90 ± 0.79	62.82 ± 2.10	65.74 ± 4.07	62.87 ± 2.64	67.38 ± 1.60	53.04 ± 2.61	67.55 ± 1.92
Average	74.08 ± 1.61	72.71 ± 1.50	71.75 ± 1.29	72.62 ± 1.83	70.62 ± 1.74	71.59 ± 2.05	69.82 ± 1.62	74.24 ± 1.71

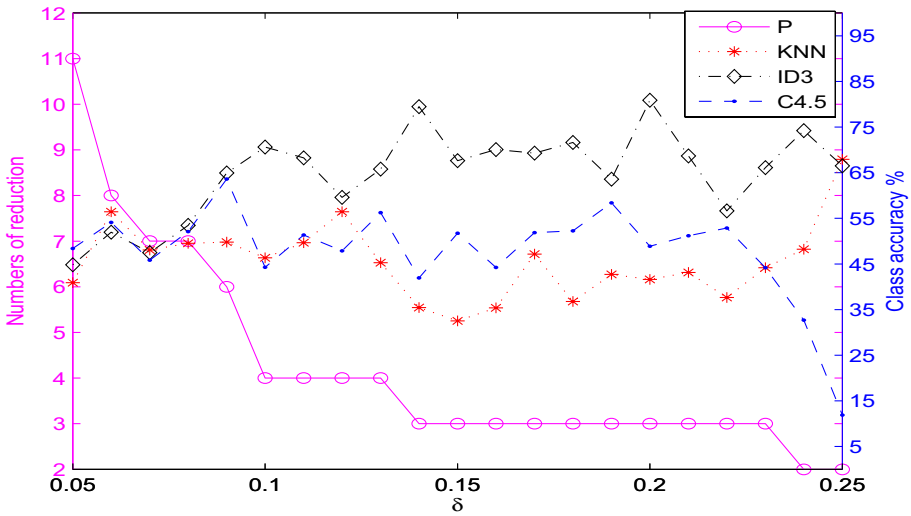


Fig. 4 Effective of δ in attribute reduction (Lung Cancer)

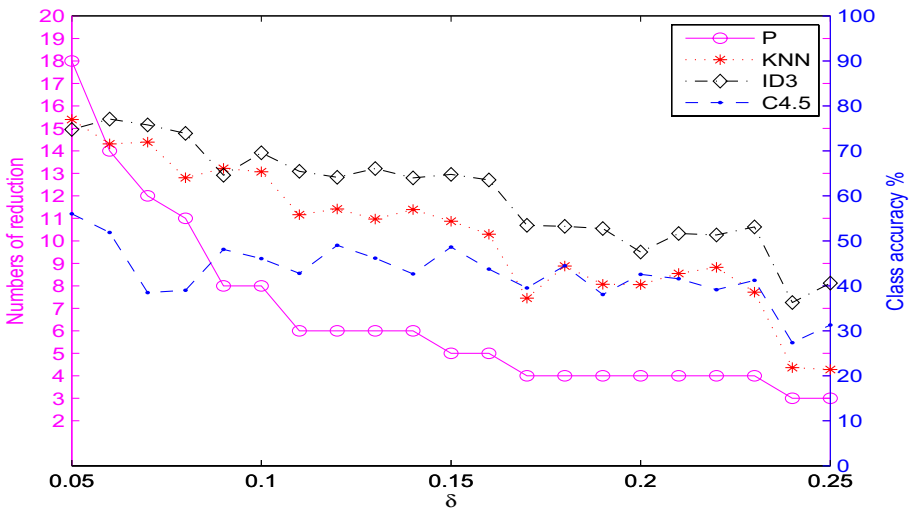


Fig. 5 Effective of δ in attribute reduction (Large Soybean)

than KNN and ID3 in most data sets, such as Large Soybean, Dermatology, Breast Cancer, Mammographic and Audiology.standardiz. This may be due to the insufficient number of samples. Because the small number of samples and the large number of total categories in Audiology.standardiz, the classification accuracy of Audiology.standardiz is low. Figure 5 shows that the classification accuracy curves of KNN, ID3 and C4.5 are very close for the Vote dataset.

Figure 11 shows the optimal attribute set size of eight data reduced by seven algorithms. Figures 12, 13 and 14 show the trend of the best classification accuracy of

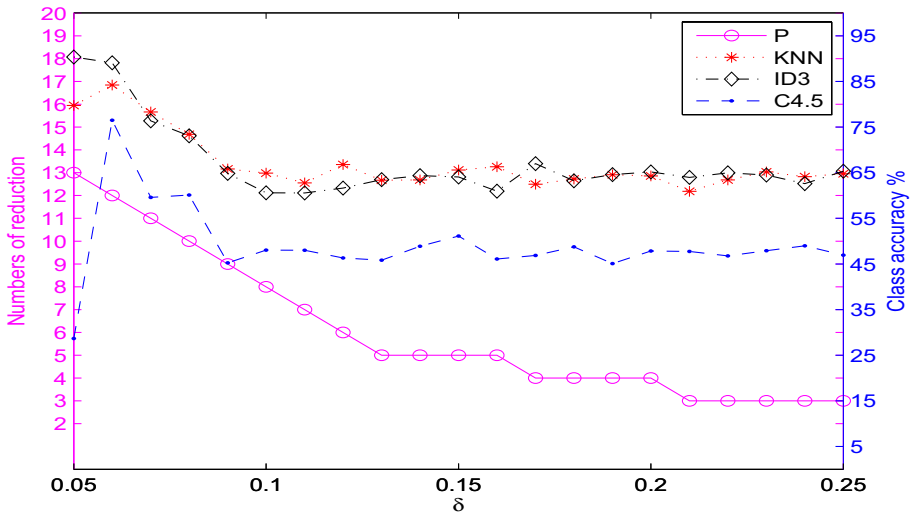


Fig. 6 Effective of δ in attribute reduction (Dermatology)

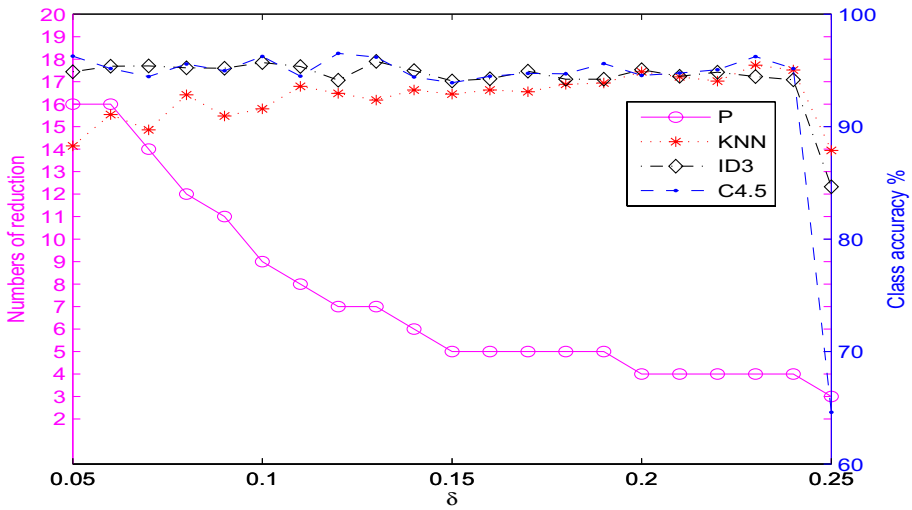


Fig. 7 Effective of δ in attribute reduction (Vote)

eight data sets under seven algorithms. The x-axis represents eight data sets, and the y-axis expresses the classification accuracy of the classifier. The pink curve, the FR-IC algorithm proposed in this paper, is obviously slightly higher than other curves. This means that the performance of FR-IC algorithm is better than other algorithms in most cases.

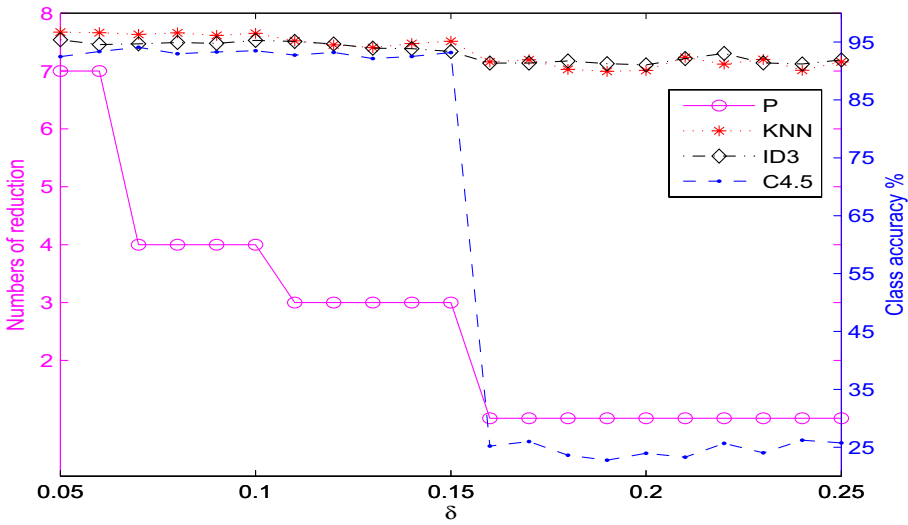


Fig. 8 Effective of δ in attribute reduction (Breast Cancer)

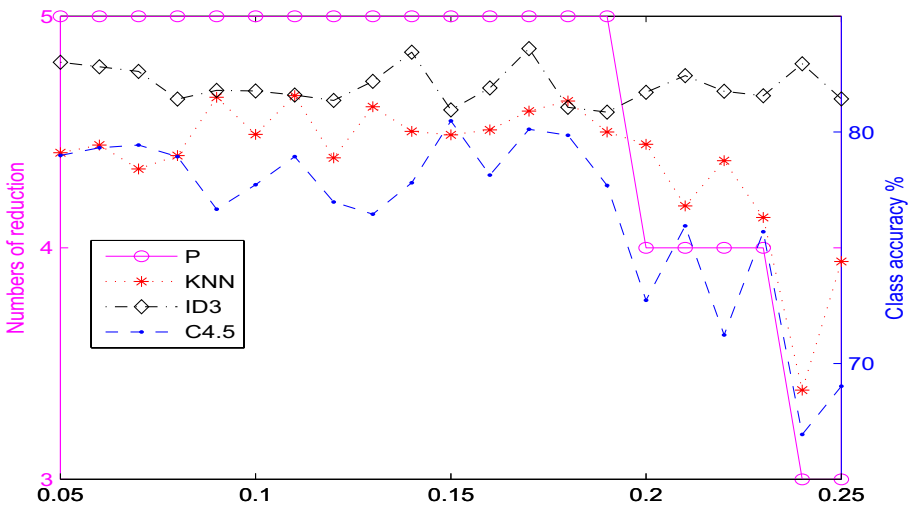


Fig. 9 Effective of δ in attribute reduction (Mammographic)

6.5 Sensitivity analysis

Next, we discuss sensitivity analysis. The original record may not be the most true because there will be deviation when observing attribute values. Two information systems Dermatology and Large Soybean are selected to generate random small disturbances on their attribute values respectively. Firstly, 0%, 2%, 5%, 8%, 15%, 20%, 25% and 30% attribute values in the information system are selected to generate random

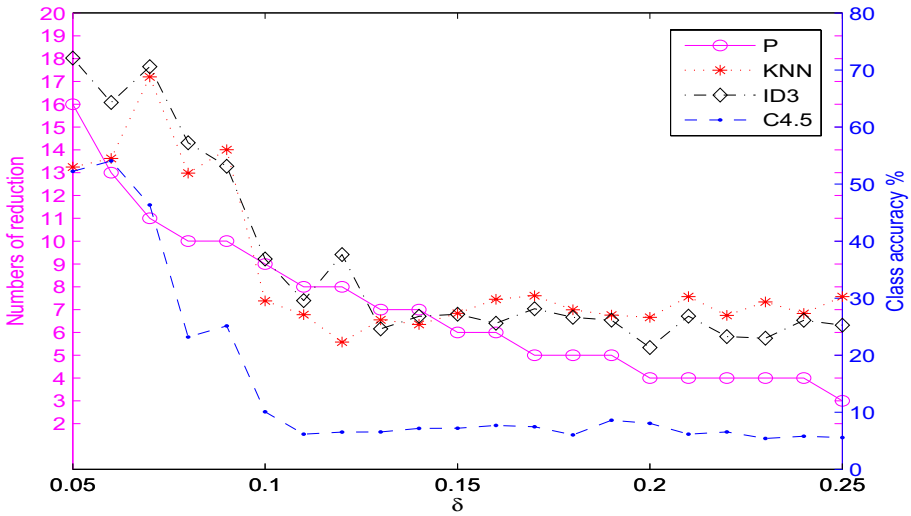


Fig. 10 Effective of δ in attribute reduction (Audiology.standardiz)

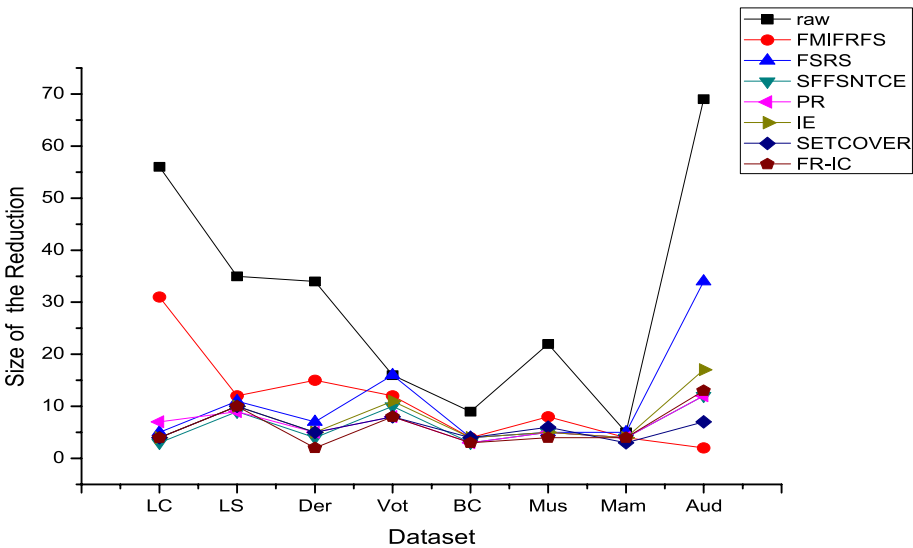


Fig. 11 The size of attribute reduction for seven algorithms

interference respectively. The purpose is to observe whether the proposed algorithm can still get a consistent reduction set under different disturbance ratios. Then the proposed algorithm is used to complete multiple attribute reduction. The difference between the reduced sets obtained under different proportions of data disturbance will be calculated. Therefore, we use Jaccard similarity coefficient to measure the similarity between the two sets:

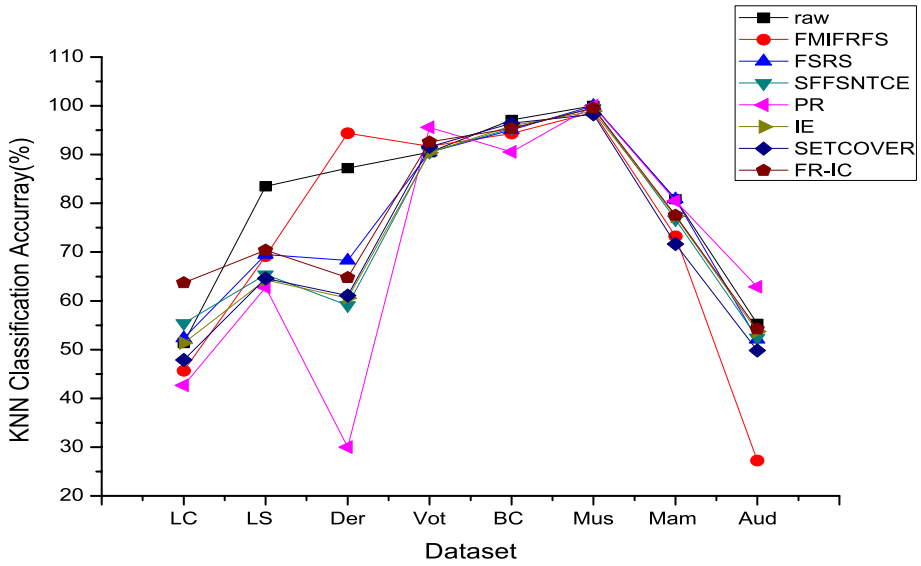


Fig. 12 Variation of classification accuracies with KNN for seven algorithms

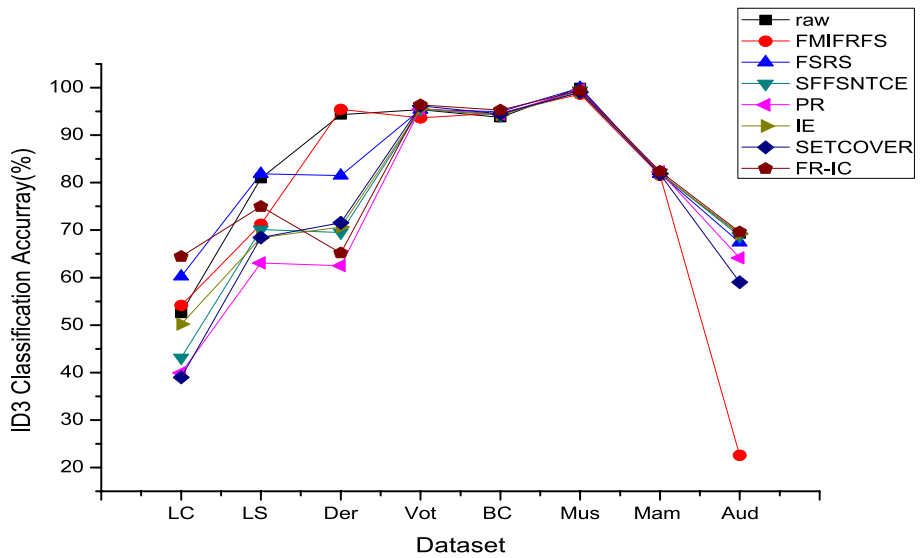


Fig. 13 Variation of classification accuracies with ID3 for seven algorithms

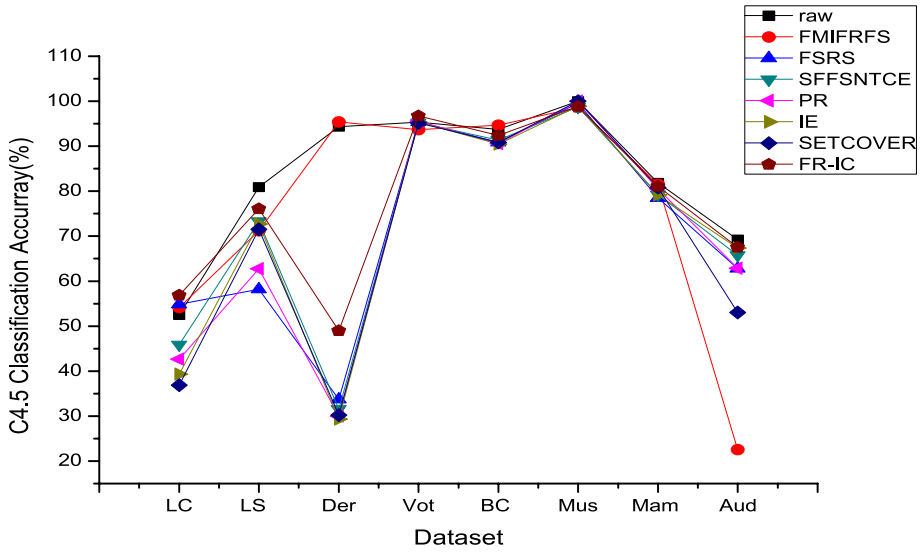


Fig. 14 Variation of classification accuracies with C4.5 for seven algorithms

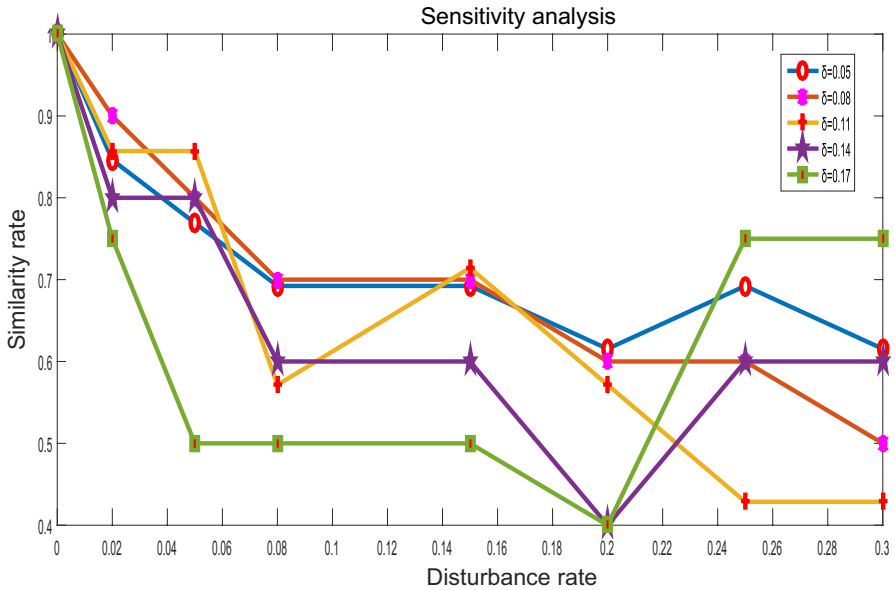


Fig. 15 Attribute sensitivity analysis (Dermatology)

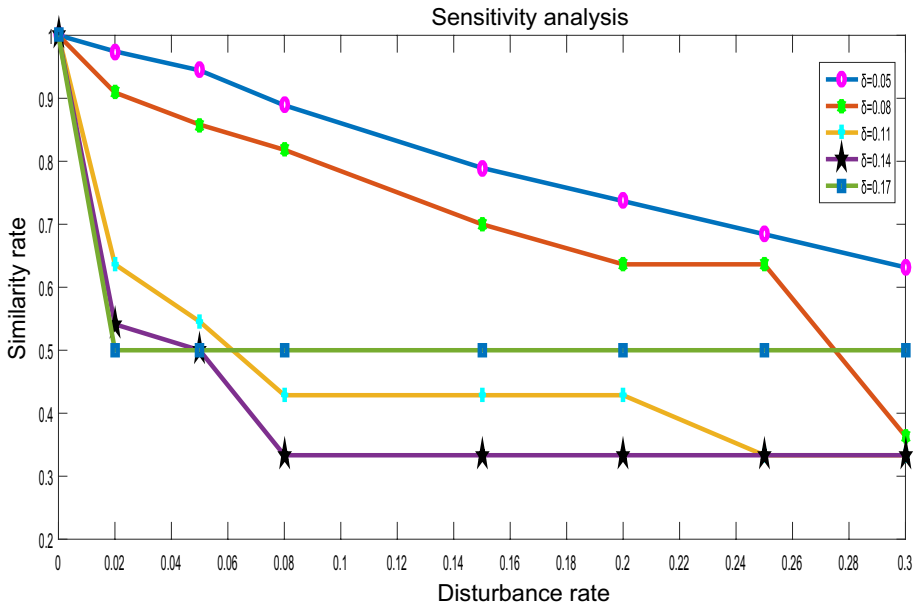


Fig. 16 Attribute sensitivity analysis (Large Soybean)

$$rate = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}$$

where $|P_i \cap P_j|$ represents the same number of statistical sets P_i and P_j , $|P_i \cup P_j|$ represents the total number of different elements in two sets. *Rate* indicates the similarity of two sets.

Figures 15 and 16 show that FR-IC algorithm obtains five different reduction sets when taking different parameters. The horizontal axis represents the disturbance rate, which is 0%, 5%, 8%, 15%, 20%, 25% and 30% respectively. When the disturbance rate is 0%, it indicates the original data set. The vertical axis represents the similarity rate. When there is no disturbance, the similarity rate is 1, and all curves start from point (0,1).

As can be seen from Figures 15 and 16, when the disturbance rate is 2%, the similarity of the reduced set is very high, reaching about 90%. When the disturbance rate increases gradually, the similarity becomes lower and lower, and finally reaches a stable state.

Through sensitivity analysis, we can know that when the disturbance rate is no more than 2%, the similarity of reduced sets is high and the algorithm model is stable.

6.6 Friedman test and Nemenyi test

To further assess the performance of classification of seven methods, Friedman test and Nemenyi test are given in this section.

Friedman test is a statistical test that uses the rank of algorithms. Friedman statistic is defined as

$$\chi_F^2 = \frac{12N}{k(k+1)} \left(\sum_{i=1}^k r_i^2 - \frac{k(k+1)^2}{4} \right)$$

where k is the number of algorithms, N is the number of data sets, r_i is the average ranking of the i -th algorithm. When k and N are large enough, Friedman statistic follows the chi-square distribution with $k - 1$ degrees of freedom. However, such Friedman test is too conservation, and is usually replaced by the next statistic

$$F_F = \frac{(N - 1)\chi_F^2}{N(k - 1) - \chi_F^2}$$

The statistic F_F follows a Fisher distribution with $k - 1$ and $(k - 1)(N - 1)$ degrees of freedom. If the statistic F_F is greater than the critical value of $F_\alpha(k - 1, (k - 1)(N - 1))$, it means the null hypothesis is rejected under Friedman test. Nemenyi test can be used to further explore which algorithm is better in the statistical term. If the average level of distance exceeds the critical distance CD_α , then the performance of two algorithms will be significantly different. The critical distance CD_α is denoted as

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}}$$

where q_α is the critical tabulated value for the test and α is the significance level of the Nemenyi test.

Below, these seven methods are viewed as seven algorithms and the statistical significance is demonstrated by using Friedman test and Nemenyi test.

- (1) The ranking of classification accuracies of the seven methods on eight data sets is given, respectively (From Tables 7, 8, 9).
- (2) Friedman test is conducted to investigate whether the classification ability of the seven methods are significantly different. Under the seven methods and the ranking of classification accuracies with three classifiers, F_F follows the distribution with 6 and 138 degrees of freedom. The critical value of Fisher distribution $F_{0.05}(6, 138)$ is 2.165, and

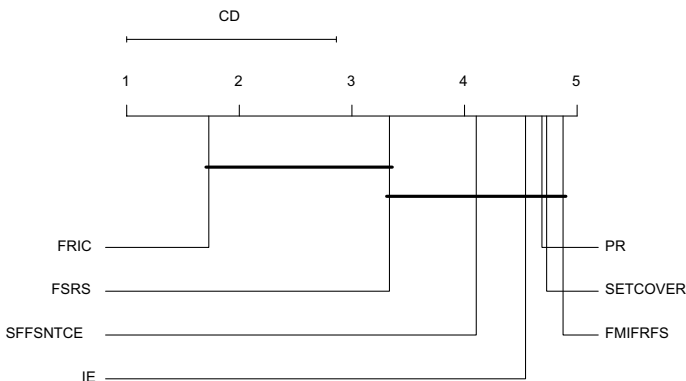


Fig. 17 The results of the Nemenyi test

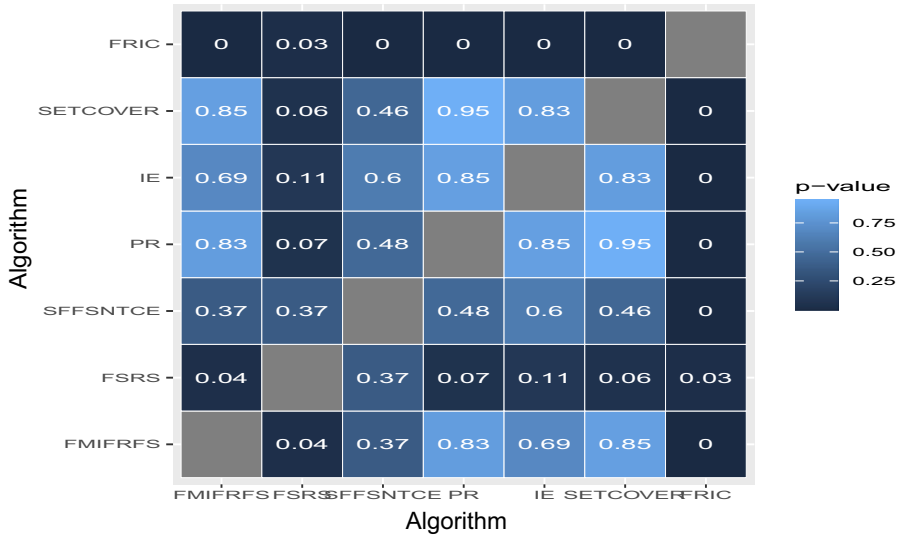


Fig. 18 Thermodynamic chart of the p -values of the Nemenyi test

Table 7 Ranking of classification accuracies of reduced data with KNN

Data sets	FMIFRFS	FSRS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
LC	7	3	2	6	4	5	1
LS	3	2	4	7	6	5	1
Der	1	2	6	7	5	4	3
Vot	2	6	5	4	7	3	1
BC	6	5	4	7	3	2	1
Mus	6	1	3	4	5	7	2
Mam	6	1	5	4	3	7	2
Aud	7	5	3	4	2	6	1
Average	4.73	3.13	4	5.38	4.38	4.88	1.5

Table 8 Ranking of classification accuracies of reduced data with ID3

Data sets	FMIFRFS	FSRS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
LC	3	2	5	6	4	7	1
LS	3	1	4	7	6	5	2
Der	1	2	5	7	4	3	6
Vot	7	6	5	3	4	2	1
BC	3	2	7	5	6	4	1
Mus	7	1	3.5	2	6	5	3.5
Mam	7	6	3	2	4	5	1
Aud	7	4	3	5	2	6	1
Average	4.75	2.63	4.44	4.63	4.5	4.63	2.06

Table 9 Ranking of classification accuracies of reduced data with C4.5

Data sets	FMIFRFS	FSRS	SFFSNTCE	PR	IE	SETCOVER	FR-IC
LC	3	2	4	5	6	7	1
LS	7	6	2	5	3	4	1
Der	1	3	4	6	7	5	2
Vot	7	3	5	2	4	6	1
BC	2	4	3	6	7	5	1
Mus	6	3	7	1.5	4	1.5	5
Mam	7	6	4	3	5	2	1
Aud	7	5	3	4	2	6	1
Average	5	4	4	4.06	4.75	4.56	1.63

the test statistic of $F_F = 9.143$. Obviously, the statistic value F_F is bigger than that of $F_{0.05}(6, 138)$. This means that at the significant level $\alpha = 0.05$, it is evidence to reject the null hypothesis, which means that the classification ability of the seven methods are different in the statistical significance.

- (3) To further show the significant difference of the seven methods, Nemenyi test is introduced. For $\alpha = 0.05$, it is easy to calculate $q_\alpha = 2.948$ and $CD_\alpha = 2.948 \times \sqrt{\frac{7 \times (7+1)}{6 \times 24}} = 1.838$. Figure 17 shows the Nemenyi test results with $\alpha = 0.05$ on the seven methods, in which the two algorithms with horizontal line connections have no significant differences on classification accuracy and the line segments in the figure curves out the scope of CD_α . Figure 18 shows the significance level p -values of each pair of algorithms.
- (4) From Figs. 17 and 18, the following results are obtained:
 - (a) The classification accuracy of FR-IC is statistically higher than that of FMIFRFS, SFFSNTCE, PR, IE and SETCOVER, respectively;
 - (b) The classification accuracy of FSRS is statistically higher than that of FMIFRFS, SFFSNTCE, PR, IE and SETCOVER, respectively;
 - (c) There is no significant difference between the classification accuracy of FR-IC and FSRS;
 - (d) There is no significant difference among the classification accuracy of FMIFRFS, SFFSNTCE, PR, IE and SETCOVER.

7 Conclusion and future work

In this paper, a fuzzy rough set model for an ICDIS has been given. Some attribute-evaluation functions, such as fuzzy positive regions, dependency functions and attribute importance functions have been presented. These attribute-evaluation functions represent the classification ability of attribute reduction. A fuzzy rough computation model for an ICDIS has been established by using the iterative relations of fuzzy positive regions and dependency functions. Attribute reduction in an ICDIS based on fuzzy rough sets has been studied and the corresponding algorithm has been proposed. Experiments have been carried out

by using 8 datasets from UCI, and statistical tests have been used to evaluate the performance of the proposed algorithm. Experimental results show that the proposed algorithm can effectively reduce redundant attributes and maintain high classification accuracy. In the future, some applications of the fuzzy rough computation model in data mining and classification learning will be studied.

Acknowledgements The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions, which have helped immensely in improving the quality of the paper. This work is supported by National Natural Science Foundation of China (11971420), Natural Science Foundation of Guangxi (AD19245102, 2020GXNSFAA159155, 2018GXNSFDA294003), Key Laboratory of Software Engineering in Guangxi University for Nationalities (2021-18XJSY-03) and Special Scientific Research Project of Young Innovative Talents in Guangxi (2019AC20052).

References

- Cattaneo G, Chiaselotti G, Ciucci D, Gentile T (2016) On the connection of hypergraph theory with formal concept analysis and rough set theory. *Inf Sci* 330:342–357
- Chen SL, Li JG, Wang XG (2005) Fuzzy sets theory and its application. Chinese Scientific Publishers, Beijing
- Chen DG, Zhang L, Zhao SY, Hu QH, Zhu PF (2012) A novel algorithm for finding reducts with fuzzy rough sets. *IEEE Trans Fuzzy Syst* 20:385–389
- Chen YM, Xue Y, Ma Y, Xu FF (2017) Measures of uncertainty for neighborhood rough sets. *Knowl Based Syst* 120:226–235
- Chen LL, Chen DG, Wang H (2019) Fuzzy kernel alignment with application to attribute reduction of heterogeneous data. *IEEE Trans Fuzzy Syst* 27(7):1469–1478
- Cornelis C, Jensen R, Martin GH, Slezak D (2010) Attribute selection with fuzzy decision reducts. *Inf Sci* 180:209–224
- Dai JH, Wang WT, Tian HW, Liu L (2013) Attribute selection based on a new conditional entropy for incomplete decision systems. *Knowl Based Syst* 39:207–213
- Dai JH, Hu H, Wu WZ, Qian YH, Huang DB (2018) Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets. *IEEE Trans Fuzzy Syst* 26(4):2175–2187
- Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. *Int J Gen Syst* 17:191–208
- Duntsch I, Gediga G (1998) Uncertainty measures of rough set prediction. *Artif Intell* 106:109–137
- Frank A, Asuncion A (2010) UCI machine learning repository [Online]. <http://archive.ics.uci.edu/ml>
- Giang NL, Son LH, Ngan TT, Tuan TM, Phuong HT, Abdel-Basset M, de Macêdo ARL, de Albuquerque VHC (2020) Novel incremental algorithms for attribute reduction from dynamic decision tables using hybrid filter-wrapper with fuzzy partition distance. *IEEE Trans Fuzzy Syst* 28(5):858–873
- Hu Q, Yu D, Pedrycz W, Chen D (2011) Kernelized fuzzy rough sets and their applications. *IEEE Trans Knowl Data Eng* 23:1649–1667
- Huang YY, Li TR, Luo C, Fujita H, Horng SJ (2017) Dynamic variable precision rough set approach for probabilistic set-valued information systems. *Knowl Based Syst* 122:131–147
- Lang GM, Miao DQ, Cai MJ (2017) Three-way decision approaches to conflict analysis using decision-theoretic rough set theory. *Inf Sci* 406–407:185–207
- Li H, Li DY, Zhai YH, Wang S, Zhang J (2016) A novel attribute reduction approach for multi-label data based on rough set theory. *Inf Sci* 367–368:827–847
- Li ZW, Zhang PF, Ge X, Xie NX, Zhang GQ, Wen CF (2019) Uncertainty measurement for a fuzzy relation information system. *IEEE Trans Fuzzy Syst* 27(12):2338–2352
- Li ZW, Huang D, Liu XF, Xie NX, Zhang GQ (2020) Information structures in a covering information system. *Inf Sci* 507:449–471
- Li ZW, Liu XF, Dai JH, Chen JL, Fujita H (2020) Measures of uncertainty based on Gaussian kernel for a fully fuzzy information system. *Knowl Based Syst* 196:105791
- Liang JY, Mi J, Wei W, Wang F (2013) An accelerator for attribute reduction based on perspective of objects and attributes. *Knowl Based Syst* 44:90–100
- Liang JY, Wang F, Dang CY (2014) A group incremental approach to feature selection applying rough set technique. *IEEE Trans Knowl Data Eng* 26:294–304
- Liu GL, Feng YB, Yang JT (2020) A common attribute reduction form for information systems. *Knowl Based Syst* 193:105466

- Meng ZQ, Shi ZZ (2009) A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets. *Inf Sci* 179:2774–2793
- Moresi NN, Yankout MM (1998) Axiomatic for fuzzy rough sets. *Fuzzy Sets Syst* 100:327–342
- Pankhuri J, Kumar TA, Tanmoy S (2020) A fitting model based intuitionistic fuzzy rough feature selection. *Eng Appl Artif Intell* 89:103421
- Pawlak Z (1982) Rough sets. *Int J Comput Inf Sci* 11:341–356
- Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht
- Radzikowska AM, Kerre EE (2002) A comparative study of fuzzy rough sets. *Fuzzy Sets Syst* 126:137–155
- Singh S, Shreevastava S, Som T, Somani G (2020) A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems. *Soft Comput* 24(6):4675–4691
- Slezak D (2002) Approximate entropy reducts. *Fundam Inf* 53:365–390
- Tan AH, Wu WZ, Tao YZ (2018) A unified framework for characterizing rough sets with evidence theory in various approximation spaces. *Inf Sci* 454–455:144–160
- Teng SH, Zhou SL, Sun JX, Li ZY (2010) Attribute reduction algorithm based on conditional entropy under incomplete information system. *J Natl Univ Def Technol* 32(1):90–94
- Tran AD, Arch-int S, Arch-int N (2018) A rough set approach for approximating differential dependencies. *Expert Syst Appl* 114:488–502
- Wang CZ, Qi Y, Shao MW, Hu QH, Chen DG, Qian YH, Lin YJ (2016) A fitting model for feature selection with fuzzy rough sets. *IEEE Trans Fuzzy Syst* 25:741–753
- Wang CZ, Huang Y, Shao MW, Fan XD (2019) Fuzzy rough set-based attribute reduction using distance measures. *Knowl Based Syst* 164:205–212
- Wang CZ, Wang Y, Shao MW, Qian YH, Chen DG (2020) Fuzzy rough attribute reduction for categorical data. *IEEE Trans Fuzzy Syst* 28(5):818–830
- Xie NX, Liu M, Li ZW, Zhang GQ (2019) New measures of uncertainty for an interval-valued information system. *Inf Sci* 470:156–174
- Yao YY (2008) Probabilistic rough set approximations. *Int J Approx Reason* 49:255–271
- Yao YY, Zhang XY (2017) Class-specific attribute reducts in rough set theory. *Inf Sci* 418–419:601–618
- Yu B, Guo LK, Li QG (2019) A characterization of novel rough fuzzy sets of information systems and their application in decision making. *Expert Syst Appl* 122:253–261
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zhang XX, Chen DG, Tsang EC (2017) Generalized dominance rough set models for the dominance intuitionistic fuzzy information systems. *Inf Sci* 378:1–25
- Zhang GQ, Li ZW, Wu WZ, Liu XF, Xie NX (2018) Information structures and uncertainty measures in a fully fuzzy information system. *Int J Approx Reason* 101:119–149
- Zhao H, Qin KY (2014) Mixed feature selection in incomplete decision table. *Knowl Based Syst* 57:181–190
- Ziarko W (1993) Variable precision rough set model. *J Comput Syst Sci* 46:39–59

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

Authors and Affiliations

Jiali He¹ · Liangdong Qu²  · Zhihong Wang³ · Yiyi Chen⁴ · Damei Luo⁵ · Ching-Feng Wen^{6,7}

Jiali He
jialihe100@126.com

Zhihong Wang
wangzhihong100@126.com

Yiyi Chen
chenyiyi2023@126.com

Damei Luo
luodamei166@126.com

- ¹ Key Laboratory of Complex System Optimization and Big Data Processing in Department of Guangxi Education, Yulin Normal University, Yulin 537000, Guangxi, People's Republic of China
- ² School of Artificial Intelligence, Guangxi University for Nationalities, Nanning 530006, Guangxi, People's Republic of China
- ³ Institute of Artificial Intelligence, School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, Sichuan, People's Republic of China
- ⁴ School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, Fujian, People's Republic of China
- ⁵ School of Mathematics and Information Science, Guangxi University, Nanning 530004, Guangxi, People's Republic of China
- ⁶ Center for Fundamental Science, and Research Center for Nonlinear Analysis and Optimization, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
- ⁷ Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung 80708, Taiwan