Maximal Discernibility Pairs Based Approach to Attribute Reduction in Fuzzy Rough Sets

Jianhua Dai, Hu Hu, Wei-Zhi Wu, Yuhua Qian, Debiao Huang

Abstract—Attribute reduction is one of the biggest challenges encountered in computational intelligence, data mining, pattern recognition and machine learning. Effective in feature selection as the rough set theory is, it can only handle symbolic attributes. In order to overcome this drawback, proposed is the fuzzy rough sets, an extended model of rough sets, which is able to deal with imprecision and uncertainty in both symbolic and numerical attributes. Existing attribute selection algorithms based on fuzzy rough set model mainly take the angle of “attribute set”, which means they define the object function representing the predictive ability for attributes subset with regard to the domain of discourse, rather than follow the view of “object pair”. Algorithms from the viewpoint of object pair can ignore the object pairs that are already discerned by the selected attributes subsets and thus need only to deal with part of object pairs instead of the whole object pairs from the discourse, which makes such algorithms more efficient in attribute selection. In this paper, we propose the concept of Reduced Maximal Discernibility Pairs, which directly adopts the perspective of object pair in the framework of fuzzy rough set model. Then we develop two attribute selection algorithms, named as Reduced Maximal Discernibility Pairs Selection (RMDPS) and Weighted Reduced Maximal Discernibility Pairs Selection (WRMDPS), based on the reduced maximal discernibility pairs. Experiment results show that the proposed algorithms are effective and efficient in attribute selection.

Index Terms—Fuzzy rough sets, attribute reduction, fuzzy discernibility matrix, maximal discernibility pairs

1 INTRODUCTION

The rough set theory, introduced by Pawlak [1], is an useful mathematical approach to deal with vague and uncertain information, has attracted many researchers’ attention and been proven to be successful in solving a variety of problems [1], [2], [3], [4], [5], [6]. However, just as [7] mentioned, rough set model can only deal with symbolic value while values of attributes might be either symbolic or real-valued in many real data sets. To overcome this problem, several extended models have been proposed [8], and fuzzy rough set model [9], [10] is a typical case. Attribute reduction, called attribute selection or feature selection as well, is regarded as one of the most important topics in rough set theory [11], [12], [13], [14], [3], [15], [16], which should be taken as a necessary preprocessing step to find a suitable subset of attributes.

In fuzzy rough set model, we use fuzzy similarity relation to replace the equivalence relation in crisp rough set theory to measure the indiscernibility between two objects. In most cases, we use a number whose value is in the unit interval to represent the degree of indiscernibility of two objects, in which 1 means they are indiscernible and 0 means they are discernable. The existing researches on fuzzy rough set contain at least two topics: the construction of the approximations of fuzzy rough set model and the applications of fuzzy rough set model. On one hand, since fuzzy rough set is proposed in [9], many different lower approximations and upper approximations have been put forward [17], more detailed information can be found in [18]. On the other hand, fuzzy rough set model has been successfully applied in many applications [19], such as classification [20], clustering [21], rule extraction [22], especially attribute reduction [3], [23], [24], [25], [26], [27].

Attribute reduction of fuzzy rough set theory has been a popular topic in recent years. Shen et al. [28] generalized the dependency function defined in classical rough set based on positive region into the fuzzy case and presented a fuzzy rough attribute selection algorithm based on such dependency. In [29], Bhatt et el. proposed an algorithm to improve its computational efficiency. In [30], a method based on fuzzy entropy was proposed. In [31], Hu et al. extended the Shannon entropy to measure the information quantity by fuzzy equivalence classes in fuzzy sets and reduce hybrid data sets. In [32], [33], fuzzy discernibility matrix was proposed, and the algorithm based on dependency [28] was improved. In [34], a novel algorithm was proposed to find reducts based on the minimum elements in a discernibility matrix and thus improved the computational efficiency of the discernibility matrix. In [35], a concept of fuzzy similarity measure and a model for evaluating
feature dependency were presented. In [36], an accelerator, called forward approximation, was constructed by combining sample reduction and dimensionality reduction together.

In summary, existing attribute reduction methods based on the fuzzy rough set model mainly adopt the angle of “attribute set”, which means they use object function to represent the predictive ability of attributes subset with regard to the domain of discourse, rather than the view of “object pair”. Comparatively speaking, algorithms based on “object pair” can ignore the object pairs that are already discerned by the selected attribute subset and thus only deal with part of object pairs each time, rather than the whole object pairs from the whole domain of discourse, which makes such algorithms more efficient. Recently, Chen et al. proposed a related algorithm (denoted by SPS) in [34], however, it is an algorithm based on crisp discernibility matrix generated by cut set technology, rather than fuzzy discernibility matrix. In essence, the framework of fuzzy rough sets was transformed into that of crisp rough sets in [34]. In this paper, we propose the concept of reduced maximal discernibility pairs. Consequently, the view of “object pair” is directly introduced into the framework of fuzzy rough sets. Then we develop two attribute selection algorithms, named as Reduced Maximal Discernibility Pairs Selection (RMIPS) and Weighted Reduced Maximal Discernibility Pairs Selection (WRMIPS), based on the reduced maximal discernibility pairs. Experiments indicate that our algorithms are effective and efficient.

The rest of this paper is organized as follows. Some related basic notions are presented in Section 2. The related definitions and concepts of Reduced Maximal Discernibility Pairs are proposed in Section 3. The proposed attribute significance measure and attribute reduction algorithms based on Reduced Maximal Discernibility Pairs are introduced in Section 4. Experiments are conducted in Section 5 and Section 6 concludes this paper.

2 Preliminaries

In this section we briefly review some basic notions about rough sets [1], [2], [33] and the attribute selection method based on rough set theory. Then, we recall the concepts of fuzzy rough sets and its corresponding attribute selection method.

2.1 Rough Sets

A decision table is defined as: \( S = \langle U, A, V, f \rangle \), where \( U = \{x_1, x_2, \ldots, x_n\} \) is a finite nonempty set of objects; \( A = C \cup D \) is a finite nonempty set of attributes, where \( C = \{c_1, c_2, \ldots, c_m\} \) is a nonempty set of conditional attributes, and \( D \) is a nonempty set of decision attributes (usually, \( D = \{d\} \)), \( C \cap D = \emptyset \), \( V \) is the union of the value domains, i.e., \( V = \bigcup_{a \in A} V_a \), where \( V_a \) is the value set of attribute \( a \), called the value domain for attribute \( a \); and \( f : U \times A \rightarrow V \) is an information function, which maps an object in \( U \) to exactly one value from domains of attribute such as \( \forall a \in A, x \in U, f(a, x) \in V_a \), where \( f(a, x) \) represents the value of object \( x \) on attribute \( a \).

Given a decision table \( S = \langle U, A, V, f \rangle \), for any subset of attributes \( B \subseteq A \), the indiscernibility relation generated by \( B \) on \( U \) is defined by

\[
IND(B) = \{(x, y) \in U^2 | f(b, x) = f(b, y), \forall b \in B\}
\]  

It is clear that \( IND(B) \) is an equivalence relation, which is reflexive, symmetric and transitive. And it determines a partition of \( U \), denoted by \( U/IND(B) \) or simply \( U/B \); an equivalence class of \( IND(B) \) containing \( x \) will be denoted by \( [x]_B \).

For any \( X \subseteq U \), the lower and upper approximations of \( X \) with respect to \( B \) can be further defined as:

\[
\underline{R}_B X = \{x | [x]_B \subseteq X\}
\]  

\[
\overline{R}_B X = \{x | [x]_B \cap X \neq \emptyset\}
\]  

\( \underline{R}_B X \) and \( \overline{R}_B X \) are two key concepts in rough set theory. Suppose \( P \) and \( Q \) are equivalence relations over \( U \), then the concepts of positive, negative and boundary regions are constructed based on lower and upper approximations as follows:

\[
POS_P(Q) = \bigcup_{X \in U/Q} \underline{R}_P X
\]  

\[
NEG_P(Q) = U - \bigcup_{X \in U/Q} \overline{R}_P X
\]  

\[
BND_P(Q) = \bigcup_{X \in U/Q} \overline{R}_P X - \bigcup_{X \in U/Q} \underline{R}_P X
\]

According to the above definitions, we find that the positive region is the collection of objects that can be discerned by attributes \( P \) with respect to decision attributes \( Q \). The negative region is the collection of objects that cannot be discerned by the given attributes \( P \) with respect to \( Q \). The boundary region is the collection of objects that might be discerned by attributes \( P \) with respect to attributes \( Q \).

Given a decision table \( S = \langle U, C \cup D \rangle \), a subset \( B \subseteq C \) is called a relative reduct of \( C \) if \( B \) is independent in \( S \), and \( POS_B(D) = POS_C(D) \). The set of all indispensable attributes in \( C \) is called the core and denoted by \( Core_C(C \cup D) \). The set of all reducts is denoted as \( Red_{\cup}(C \cup D) \), and we have \( Core_C(C \cup D) = \cap Red_{\cup}(C \cup D) \).

For any \( B \subseteq C \), we can say that the decision attribute \( D \) depends on \( B \) to degree \( \gamma_B(D) \), defined as follows:

\[
\gamma_B(D) = \frac{|POS_B(D)|}{|U|}
\]  

\( \gamma_B(D) = 1 \) means \( D \) totally depends on \( B \); \( D \) partially depends on \( B \) when \( 0 < \gamma_B(D) < 1 \); and when \( \gamma_B(D) = 0 \), \( D \) does not depend on \( B \) at all.

Attribute reduction based the positive region in rough set theory is to find a subset \( B \) of conditional attribute \( C \), which is a minimal set preserving the value of \( \gamma \). In other

<table>
<thead>
<tr>
<th>Table 1: A decision table</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U )</td>
</tr>
<tr>
<td>( x_1 )</td>
</tr>
<tr>
<td>( x_2 )</td>
</tr>
<tr>
<td>( x_3 )</td>
</tr>
<tr>
<td>( x_4 )</td>
</tr>
</tbody>
</table>
words, a subset of conditional attributes can be regarded as a reduct only if it satisfies:

1) $\gamma_B(D) = \gamma_C(D)$  \hspace{1cm} (8)
2) $\forall B' \subset B, \gamma_B'(D) < \gamma_B(D)$  \hspace{1cm} (9)

There might exist many reducts for a data set. The discernibility matrix, introduced by Skowron and Rauszer [37] to find reducts based on the rough set theory, is a matrix in which conditional attributes that can discern pairs of objects are stored. In other words, it can be denoted by $M = (M(x, y))$ which is a $|U| \times |U|$ matrix, and each of its entry for a given decision table $S = < U, C \cup D >$ is defined by

$$M(x, y) = \{a \in C | f(a, x) \neq f(a, y) \text{ and } f(D, x) \neq f(D, y)\}$$  \hspace{1cm} (10)

The implication of matrix entry $M(x, y)$ is that any object pair $(x, y)$ can be differentiated by the attributes in $M(x, y)$, which characterize the ability of object pair $(x, y)$. A discernibility matrix $M$ is symmetric, i.e., $M(x, y) = M(y, x)$, and $M(x, x) = \emptyset$. Therefore, it is sufficient to consider only the lower triangle or the upper triangle of the matrix.

Given a decision table $S = < U, C \cup D >$, an attribute $c_i \in C$ belongs to core $Core_U(C \cup D)$ iff $3x, y \in U$ satisfying $M(x, y) = \{c_i\}$. Thus in discernibility matrix, the core is defined by:

$$Core_U(C \cup D) = \bigcup \{M(x, y)|x, y \in U \text{ and } |M(x, y)| = 1\}$$  \hspace{1cm} (11)

where $|M(x, y)|$ means the number of attributes contained in $M(x, y)$.

The discernibility function $f_D$ as one of the key concepts of the discernibility matrix, is a Boolean function which can be defined as follows:

$$f_D(c_1, c_2, ..., c_m) = \wedge \{\forall x, y \in U, \text{ and } |M(x, y)| > 0\}$$  \hspace{1cm} (12)

Thus for discernibility matrix, any reduct for a decision table is the set of attributes $B \subseteq C$ satisfying:

1) $\forall x, y \in U, M(x, y) \neq \emptyset \rightarrow B \cap M(x, y) \neq \emptyset$  \hspace{1cm} (13)
2) $\forall B' \subset B, 3x, y \in U, M(x, y) \neq \emptyset \rightarrow B' \cap M(x, y) = \emptyset$  \hspace{1cm} (14)

Through the discernibility matrix, we are able to find one or all reducts in a given data set.

### 2.2 Fuzzy Rough Sets

The traditional rough set theory can only process symbolic-valued attributes. However, most of real data sets contain real-valued attributes, which means it goes beyond the capacity of the traditional rough set theory. Hence, the crisp rough set model is extended the fuzzy rough set model. By means of fuzzy rough set model [38], [39], we can handle the real-valued attributes or the hybrid attributes directly.

In fuzzy rough sets, we use fuzzy similarity relation $R$ to substitute the crisp equivalence relation in traditional rough sets. A fuzzy similarity relation $R$ is a fuzzy relation $S : U \times U \rightarrow [0, 1]$ which should satisfy the following conditions:

- Reflexivity: $\forall x \in U, S(x, x) = 1$
- Symmetry: $\forall x, y \in U, S(x, y) = S(y, x)$

- T-transitivity: $\forall x, y, z \in U, S(x, z) \geq T(S(x, y), S(y, z))$

Here, $T$ is a T-norm [17] which is an associative aggregation operator $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ holds the following conditions:

- commutativity: $T(a, b) = T(b, a)$.
- monotonicity: $T(a, b) \leq T(c, d)$, if $a \leq c$ and $b \leq d$.
- associativity: $T(a, T(b, c)) = T(T(a, b), c)$.
- boundary conditions: $T(a, 1) = a$.

$I : [0, 1]^2 \rightarrow [0, 1]$ is an implicator [40], which satisfies $I(0, 0) = 1$, $I(1, 0) = 0$ and $I(0, 1) = 1$. An implicator $I$ is called left monotonic iff $I(., x)$ decreases $\forall x \in [0, 1]$. Similarly, $I$ is called right monotonic iff $I(x, .)$ increases $\forall x \in [0, 1]$. Once $I$ is both left and right monotonic, then it can be called as hybrid monotonic.

In [40], Radzikowska et al. proposed the lower and upper approximations by means of T-transitive fuzzy similarity relation:

$$\mu_{R_p} X(x) = \inf_{y \in U} I(\mu_{R_p}(x, y), \mu_X(y))$$  \hspace{1cm} (15)
$$\mu_{T_p} X(x) = \sup_{y \in U} T(\mu_{R_p}(x, y), \mu_X(y))$$  \hspace{1cm} (16)

Where, $I$ and $T$ mean fuzzy implicator and T-norm respectively, and $R_p$ refers to the similarity relation induced by the subset of attributes $P$

$$\mu_{R_p}(x, y) = \min_{a \in P} \{\mu_{R_a}(x, y)\}$$  \hspace{1cm} (17)

Where, $\mu_{R_a}(x, y)$ is the similarity degree of objects $x$ and $y$ with respect to attribute $a$.

In [7], the fuzzy positive region is proposed by means of extension principle [41] according to the crisp positive region in the traditional rough set theory. Thus $\forall x \in U$, its membership with respect to the positive region can be defined as follows:

$$\mu_{POS^Q}(x) = \sup_{X \in U/Q} \mu_{R_p} X(x)$$  \hspace{1cm} (18)

The fuzzy rough dependency is defined by

$$\gamma_{\mu} (Q) = \frac{\sum_{x \in U} \mu_{POS^Q}(x)}{|U|}$$  \hspace{1cm} (19)

With fuzzy dependency, we are able to measure the ability of a given subset of attributes for preserving the dependency degree of the entire attributes. By means of comparing fuzzy dependency, we can find a way to choose an attribute subset $B$ which can provide the same predictive ability as $C$, i.e. $\gamma_B(D) = \gamma_C(D)$. In other words, a subset of conditional attributes can be regarded as a reduct only if it satisfies:

1) $\gamma_B(D) = \gamma_C(D)$  \hspace{1cm} (20)
2) $\forall B' \subset B, \gamma_B'(D) < \gamma_B(D)$  \hspace{1cm} (21)

Similar to the crisp case, the discernibility matrix in the fuzzy rough set model is also an important approach to find one or more reducts. The fuzzy discernibility matrix, proposed in [33], is an extension of the crisp discernibility matrix in rough set theory. For a given decision table $S = < U, C \cup D >$, each entry of the corresponding fuzzy
discernibility matrix, denoted by $M'(x, y)$, is defined as follows:

$$M'(x, y) = \{a|N(\mu_{R_a}(x, y))|a \in C \} \quad \forall x, y \in U$$  \hspace{1cm} (22)

Where, $\mu_{R_a}(x, y)$ describes the fuzzy indiscernibility degree between objects $x$ and $y$ with respect to attribute $a$. On the contrary, $N(\mu_{R_a}(x, y)) = 1 - \mu_{R_a}(x, y)$ means the fuzzy discernibility degree among objects $x$ and $y$ with regard to attribute $a$. For example, an entry $M'(x, y)$ might be $\{a_0, a_1, a_2\}$. The Core$_D'(C)$ in the fuzzy discernibility matrix is defined by:

$$Core_D'(C) = \{a|\exists M'(x, y), \mu_{R_a}(x, y) > 0, \forall c \in C - \{a\}, \mu_{R_c}(x, y) = 0\}$$  \hspace{1cm} (23)

Then the fuzzy discernibility function $f'_D$, by extending the concept of the discernibility matrix in rough set theory, is defined as follows:

$$f'_D(c_1, c_2, ..., c_m) = \land\{\forall M'(x, y) \leftarrow N(\mu_{R_D}(x, y))|\forall x, y \in U\}$$  \hspace{1cm} (24)

Here, $\leftarrow$ represents the fuzzy implication and $D$ means the decision attributes. It is noteworthy that, the same as discernibility function, the satisfaction of the clause is largely affected by the value of decision attributes. Since in this paper, we concentrate on the data sets which only largely affected by the value of decision attributes. Since each entry can be represented by an object pair $(x, y)$, we can also call such algorithms based on “object pair”. One important property in attribute reduction based on crisp discernibility matrix is: for a given reduct candidate $B$, any entry containing at least one attribute that belongs to $B$ can be neglected. In this paper, we extend such a property to the fuzzy discernibility matrix by means of constructing the reduced maximal discernibility pairs, which means such pair can be discerned by any attribute in it.

We first introduce two important concepts in fuzzy rough sets:

$$sim_a(x, y) = \mu_a(x, y), a \in C$$  \hspace{1cm} (29)

$$dis_a(x, y) = N(\mu_a(x, y)), a \in C$$  \hspace{1cm} (30)

$sim_a(x, y)$ represents the fuzzy indiscernibility between objects $x$ and $y$, on the contrary $dis_a(x, y)$ represents the fuzzy discernibility. For simplicity, we use $N(x) = 1 - x$ in this paper. Note that, in this paper we assume that the fuzzy similarity satisfies symmetry:

$$sim_a(x, y) = sim_a(y, x)$$  \hspace{1cm} (31)

In order to find a reduct, we follow the idea of defining the degree for a given entry $M'(x, y)$ as Eq. 25. But in this paper, we intend to use operator max to specify it:

$$SAT_{B,D}(M'(x, y)) = \max_{a \in B} \{1 - \mu_{R_a}(x, y)\} \leftarrow N(\mu_{R_D}(x, y))$$  \hspace{1cm} (32)

Which means, same as the crisp discernibility matrix, we concentrate only on the attributes having the maximal discernibility, and ignore the others. We can also express it as follows:

$$SAT_{B,D}(M'(x, y)) = \min_{a \in B} \{\mu_{R_a}(x, y)\} \leftarrow N(\mu_{R_D}(x, y))$$  \hspace{1cm} (33)

which means the minimal fuzzy similarity and maximal fuzzy discernibility are equivalent.

So far, we find that for any $M'(x, y)$, any attribute contained in it should be concerned only when the attribute has the minimal fuzzy similarity or the maximal fuzzy discernibility. Thus the minimal fuzzy similarity attributes and the maximal fuzzy discernibility attributes are defined as follows:

Definition 1. For a given decision table $S = < U, C \cup D >$, the minimal fuzzy similarity attributes with respect to pair $(x, y)$ in the fuzzy discernibility matrix is defined as follows:

$$MSA_{C,D}(x, y) = \{a|sim_a(x, y) = \min_{a \in C} sim_a(x, y), a \in C\}$$

$\leftarrow N(\mu_{R_D}(x, y))$  \hspace{1cm} (34)

Where $\leftarrow$ is a symbol representing a switch used to determine whether its left side should be calculated, i.e. $a \leftarrow b$ means: if $b = 1$, then calculate $a$; if $b = 0$, then do not calculate $a$ and assign empty set to $a$ directly. In other words, Eq. 34 is equal to the following equation:

$$MSA_{C,D}(x, y) = \begin{cases} \{a|sim_a(x, y) = \min_{a \in C} sim_a(x, y), & an \in C\} \\ \emptyset, & if \mu_{R_D}(x, y) = 0 \end{cases}$$

$\leftarrow N(\mu_{R_D}(x, y))$  \hspace{1cm} (35)
The maximal fuzzy discernibility attributes corresponding to pair \((x, y)\) in the fuzzy discernibility matrix is defined as:

\[
MDA_C D(x, y) = \{a | \text{dis}_a(x, y) = \max_{a \in C} \text{dis}_c(x, y), a \in C\} = \mathcal{P}(\mathcal{N}(\mu_{R_D}(x, y)))
\]

(36)

In this paper, we use \(|MSA_C(x, y)|\) and \(|MDA_C(x, y)|\) to represent the number of conditional attributes contained in the minimal fuzzy similarity attributes and the maximal fuzzy discernibility attributes respectively.

Eq. 36 is equal to the following equation

\[
MDA_C D(x, y) = \begin{cases}
\{a | \text{dis}_a(x, y) = \max_{c \in C} \text{dis}_c(x, y), \ a \in C\}, & \text{if } \mu_{R_D}(x, y) = 0 \\
\emptyset, & \text{if } \mu_{R_D}(x, y) = 1
\end{cases}
\]

(37)

The above definition can successfully degrade to the crisp form, in which \(MSA_C(x, y)\) and \(MDA_C(x, y)\) are the same as \(M(x, y)\) in the crisp discernibility matrix where the range is \([0, 1]\).

**Proposition 1.** For a given decision table \(S = \langle U, C \cup D\rangle\), for any pair \((x, y)\) in the corresponding fuzzy discernibility matrix, we have

\[
MDA_C D(x, y) = MDA_C D(y, x)
\]

(38)

\[
MSA_C D(x, y) = MSA_C D(y, x)
\]

(39)

\[
MDA_C D(x, y) = MSA_C D(x, y)
\]

(40)

**Proof:** Since the fuzzy discernibility matrix is symmetric. According to Definition 1, Eq. 38 and Eq. 39 can be easily obtained. The result in Eq. 40 follows directly by using Eq. 29, Eq. 30 and Definition 1.

Then, we propose the concepts of the minimal fuzzy similarity pairs and the maximal fuzzy discernibility pairs.

**Definition 2.** For a given decision table \(S = \langle U, C \cup D\rangle\), the minimal fuzzy similarity pairs with respect to attribute \(a \in C\) can be defined by:

\[
MSP_a D(U) = \{(x, y) | a \in MSA_C D(x, y), (x, y) \in U \times U\}
\]

(41)

The maximal fuzzy discernibility pairs with respect to any attribute \(a \in C\) is defined as:

\[
MDP_a D(U) = \{(x, y) | a \in MDA_C D(x, y), (x, y) \in U \times U\}
\]

(42)

In the following, we use \(|MSP_a D(U)|\) and \(|MDP_a D(U)|\) to represent the sizes of the minimal fuzzy similarity pairs and the maximal fuzzy discernibility pairs respectively.

**Proposition 2.** For a given decision table \(S = \langle U, C \cup D\rangle\), for any attribute \(a \in C\), we have

\[
MSP_a D(U) = MDP_a D(U)
\]

(43)

**Proof:** From Proposition 1 we know that \(MDA_C D(x, y) = MSA_C D(x, y)\). According to Definition 2, it is easy to get \(MSP_a D(U) = MDP_a D(U)\).

**Example 1.** Considering a decision table \(S = \langle U, C \cup D\rangle\) shown in Table 1, we use the fuzzy similarity measure defined in Eq. 54. Then the fuzzy relation is calculated as follows:

\[
sim_{c_1} = \begin{pmatrix}
1.000 & 0.580 & 0.000 & 0.000 \\
0.580 & 1.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 0.580 \\
0.000 & 0.000 & 0.580 & 1.000
\end{pmatrix}
\]

(44)

Thus:

\[
MDP_{c_1} D(U) = \{(x_1, x_1), (x_1, x_3), (x_1, x_3), (x_1, x_4)\}
\]

\[
MDP_{c_2} D(U) = \{(x_1, x_1), (x_1, x_3), (x_2, x_3), (x_3, x_4), (x_4, x_3)\}
\]

\[
MDP_{c_3} D(U) = \{(x_1, x_1), (x_1, x_4), (x_1, x_4), (x_1, x_4)\}
\]

\[
MSA_{c_1} D(U) = \{(x_1, x_1), (x_1, x_4), (x_1, x_4), (x_1, x_4)\}
\]

\[
MSA_{c_2} D(U) = \{(x_1, x_1), (x_1, x_4), (x_1, x_4), (x_1, x_4)\}
\]

\[
MSA_{c_3} D(U) = \{(x_1, x_1), (x_1, x_4), (x_1, x_4), (x_1, x_4)\}
\]

As Eqs. 43 and 40 illustrate, the maximal fuzzy discernibility pairs and the minimal fuzzy similarity pairs are the same. Hence, we concentrate only on the maximal fuzzy discernibility pairs in the following for the purpose of simplicity.

**Definition 3.** For a given decision table \(S = \langle U, C \cup D\rangle\), the maximal fuzzy discernibility pairs with respect to attribute subset \(B \subseteq C\) can be defined in the following way:

\[
MDP_B D(U) = \bigcup_{a \in B} MDP_a D(U)
\]

(44)

Here, \(MDP_B D(U)\) is a set that contains all the maximal discernibility pairs with respect to the attribute in \(B\).

**Proposition 3.** For a given decision table \(S = \langle U, C \cup D\rangle\), for any attribute subset \(B \subseteq C\), we have

\[
MDP_B D(U) \subseteq MDP_C D(U), \forall B \subseteq C
\]

(45)

**Proof:** From Definition 3, it is easy to know if \(B \subseteq C\), \(MDP_B D(U) \subseteq MDP_C D(U)\).

**Proposition 4.** For a given decision table \(S = \langle U, C \cup D\rangle\), \(\forall B' \subseteq B \subseteq C\), \(MDP_{B'} D(U) \subseteq MDP_B D(U)\).

**Proof:** According to Definition 3, it can be proved easily.

According to Proposition 4, we know that \(|MDP_B D(U)|\) satisfies monotonicity with respect to attribute subset \(B\).

**Proposition 5.** For a given decision table \(S = \langle U, C \cup D\rangle\), \(\forall B' \subseteq B \subseteq C - B\), we have

\[
|MDP_{B'} D(U)| \leq |MDP_B D(U)|
\]

(46)

\[
|MDP_B D(U)| \leq |MDP_{B \cup \{a\}} D(U)|
\]

(47)

**Proposition 6.** For a given decision table \(S = \langle U, C \cup D\rangle\) iff \(|MDP_B D(U)| = |MDP_C D(U)|\).

1063-6706 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Proposition 7. For a given decision table $S = \langle U, C \cup D \rangle$, $B$ is a reduct of the given decision table if it satisfies:

1) $MDP_B D(U) = MDP_D D(U)$
2) $\forall B' \subseteq B, MDP_{B'} D(U) \subseteq MDP_B D(U)$

Proof: According to Proposition 3, $MDP_B D(U) = MDP_D D(U)$ if and only if $|MDP_B D(U)| = |MDP_D D(U)|$. Thus the proposition holds.

Proposition 8. For a given decision table $S = \langle U, C \cup D \rangle$, $B$ is a reduct of the given decision table if $S$ is the union of $MDP_B D(U)$, $a \in B$. Hence, it is easy to get $MDP_B D(U) \subseteq MDP_B D(U)$ if $\forall B' \subseteq B$.

Proposition 9. For a given decision table $S = \langle U, C \cup D \rangle$, $MDP_B D(U) = MDP_D D(U)$ if $|MDP_B D(U)| = |MDP_D D(U)|$.

Proof: According to Definition 5 and Proposition 5, this proposition can be easily proved.

Proposition 10. For a given decision table $S = \langle U, C \cup D \rangle$, $B$ is a reduct of the given decision table if:

1) $MDP_B D(U) = MDP_4 D(U)$
2) $\forall B' \subseteq B, MDP_{B'} D(U) \subseteq MDP_B D(U)$

Proof: According to Definition 5 and Proposition 5, this proposition can be easily proved.

Example 2. Let us consider the data set in Table 1 again. Then $MDP_{c_1,c_2} D(U) = \{(x_1, x_1), (x_4, x_1), (x_2, x_3), (x_3, x_2), (x_2, x_2), (x_4, x_3), (x_3, x_3), (x_4, x_3), (x_2, x_2), (x_2, x_1)\}$

$MDP_{c_1,c_3} D(U) = \{(x_1, x_1), (x_4, x_1), (x_2, x_3), (x_2, x_2), (x_3, x_3), (x_4, x_3), (x_3, x_3), (x_2, x_3), (x_2, x_1)\}$

$MDP_{c_2,c_3} D(U) = \{(x_4, x_1), (x_2, x_3), (x_3, x_3), (x_4, x_3), (x_3, x_3), (x_2, x_3), (x_2, x_1)\}$

$MDP_{c_1,c_2,c_3} D(U) = \{(x_1, x_1), (x_4, x_1), (x_2, x_3), (x_2, x_2), (x_3, x_3), (x_4, x_3), (x_3, x_3), (x_2, x_3), (x_2, x_1)\}$

Characteristics of $MDP$ and $MDP^{'}$ indicate that we can evaluate attributes subset from the viewpoint of “object pair”. As for attribute reduction, methods from the angle of object pair can ignore the object pairs that are already discerned by the selected attributes subsets and thus need only to deal with part of object pairs instead of the whole object pairs from the discourse, which makes such algorithms efficient in attribute selection.

4 Algorithms Based on The Maximal Discernibility Pairs Selection

According to Propositions 8, 9 and 10, we find that the reduced maximal discernibility pairs are suitable for attribute selection. In this section, we propose two algorithms, called Reduction based on Maximal Discernibility Pairs Selection (denoted by RMDFS) and Weighted Reduction based on Maximal Discernibility Pairs Selection (denoted by WRMDFS), in the framework of the reduced maximal discernibility pairs.
is similar to the algorithm in [42] for crisp discernibility discernibility pairs with respect to each attribute, which ensuring the importance of attribute by the size of the maximal

Algorithm 1 RMDPS

Input: A decision table \( S = (U, C \cup D, V, f) \), where \( U = \{x_1, x_2, \ldots, x_n\}, C = \{c_1, c_2, \ldots, c_m\}, D = \{d\} \)

Output: Red
1: Red = \emptyset; maxNum = 0; pairNum[i] = 0, 1 \leq i \leq |C|.
2: Compute \( MDP_C.D(U) \);
3: while true do
4: pairNum[i] = 0, 1 \leq i \leq |C|;
5: for all \( (x_i, x_j) \in MDP_C.D(U) \) do
6: for all \( c_k \in MDA_C.D(x_i, x_j) \) do
7: pairNum[k]++;
8: end for
9: end for
10: maxNum = 0;
11: for all \( c_k \in C, 1 \leq k \leq |C| \) do
12: if pairNum[k] > maxNum then
13: maxNum = pairNum[k];
14: selAtt = \( k \);
15: end if
16: end for
17: if maxNum \neq 0 then
18: Red = Red \cup c_{\text{selAtt}};
19: \( MDP'_C.D(U) = MDP_C.D(U) - MDP'_{c_{\text{selAtt}}}.D(U) \);
20: else
21: BREAK;
22: end if
23: end while

4.1 RMDPS

Algorithm 1, denoted by RMDPS, is proposed via measuring the importance of attribute by the size of the maximal discernibility pairs with respect to each attribute, which is similar to the algorithm in [42] for crisp discernibility matrix.

In Algorithm 1, pairNum[i] represents the number of object pairs in the remaining maximal discernibility pairs with respect to attribute \( c_i \). Each time we choose the first attribute with the maximal value for pairNum as the most important attribute i.e. \( \text{selAtt} \), then we delete any object pair \( (x_i, x_j) \) that \( \text{selAtt} \in MDA_C.D(x_i, x_j) \).

Step 2 consists of two parts: calculating the fuzzy relations and getting reduced maximal discernibility pairs. As shown in Fig.1, the time complexity is \( O(|U|^2|C|) \). The time complexity from step 3 to 17 is \( O(|U|^2|C| \cdot \frac{1}{1-\delta}) \), in which \( \delta = \frac{|U|}{|U|-|V|} \). The detailed analysis can be found in Appendix. In summary, the time complexity of RMDPS is \( \max(O(|U|^2|C|), O(|U|^2|C| \cdot \frac{1}{1-\delta})) \). According to Proposition 11 in the Appendix, \( O(|U|^2|C| \cdot \frac{1}{1-\delta}) \) is smaller compared with \( O(|U|^2|C|) \) when \( \text{red} \) is small. In order to make our algorithm easier to understand, we also present it in a flow chart, shown in Fig. 2.

4.2 WRMDPS

Algorithm 2, denoted by WRMDPS, is a “weighted” version of RMDPS. It differs from RMDPS at step 6. In Algorithm 1, we directly use the size of the remaining reduced maximal discernibility pairs with respect to each attribute to measure the importance of the attribute. However, in Algorithm 2, we combine the size of the remaining reduced maximal discernibility pairs and the value of \( |MDP_C(x_i, x_j)| \) together to measure the importance of each attribute. In other words, the importance of each attribute \( c_i \) is measured by the value of \( |MDP'_C.D(U)| \) and the value of \( |MDP_C(x_i, x_j)| \) in which \( (x_i, x_j) \in MDP'_C.D(U) \).

Its time complexity is \( \max(O(|U|^2|C|), O(|U|^2|C| \cdot \frac{1}{1-\delta})) \), and the detailed analysis is similar to the time complexity in Section 4.1.
Table 2: The detailed information of the data sets

<table>
<thead>
<tr>
<th>Index</th>
<th>Data Set</th>
<th>Abbreviation</th>
<th>Objects</th>
<th>Conditional Attributes</th>
<th>Data type</th>
<th>Decision classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glass</td>
<td>glass</td>
<td>214</td>
<td>22</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>2</td>
<td>Horse Colic</td>
<td>horse</td>
<td>368</td>
<td>22</td>
<td>Nominal</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>3</td>
<td>Ecoli</td>
<td>ecoli</td>
<td>336</td>
<td>7</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>4</td>
<td>clean</td>
<td>clean</td>
<td>476</td>
<td>166</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>5</td>
<td>credit</td>
<td>credit</td>
<td>1000</td>
<td>20</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>6</td>
<td>Lymphymph</td>
<td>lymph</td>
<td>148</td>
<td>18</td>
<td>Numeric, Nominal</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>ColonAll</td>
<td>colon</td>
<td>62</td>
<td>2000</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>8</td>
<td>Hepatocellular</td>
<td>hep</td>
<td>33</td>
<td>7129</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>9</td>
<td>ionosphere</td>
<td>ionsphere</td>
<td>501</td>
<td>34</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>11</td>
<td>segment</td>
<td>segment</td>
<td>2310</td>
<td>19</td>
<td>Numeric</td>
<td>Nominal, Numeric</td>
</tr>
<tr>
<td>12</td>
<td>anneal</td>
<td>anneal</td>
<td>798</td>
<td>38</td>
<td>Nominal, Nominal</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3: Comparison of the whole running time (seconds)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>alma</th>
<th>anneal</th>
<th>clean</th>
<th>credit</th>
<th>ecoli</th>
<th>glass</th>
<th>hepa</th>
<th>hep</th>
<th>horse</th>
<th>ions</th>
<th>lymph</th>
<th>segm</th>
<th>AverageTime</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMdps</td>
<td>0.863</td>
<td>1.467</td>
<td>1.365</td>
<td>0.136</td>
<td>0.822</td>
<td>0.118</td>
<td>0.011</td>
<td>0.714</td>
<td>0.007</td>
<td>0.055</td>
<td>0.096</td>
<td>0.003</td>
<td>7.035</td>
</tr>
<tr>
<td>WRMDPS</td>
<td>235.571</td>
<td>13.73</td>
<td>44.67</td>
<td>16.479</td>
<td>6.02</td>
<td>6.908</td>
<td>16.22</td>
<td>6.09</td>
<td>6.15</td>
<td>34.82</td>
<td>1.15</td>
<td>7.035</td>
<td>1.24</td>
</tr>
<tr>
<td>L-FRPS</td>
<td>1.428</td>
<td>4.119</td>
<td>4.408</td>
<td>0.22</td>
<td>1.923</td>
<td>0.022</td>
<td>0.012</td>
<td>0.171</td>
<td>0.222</td>
<td>0.135</td>
<td>0.011</td>
<td>0.009</td>
<td>8.61</td>
</tr>
<tr>
<td>FDM</td>
<td>0.07</td>
<td>1.52</td>
<td>1.394</td>
<td>0.104</td>
<td>0.032</td>
<td>0.018</td>
<td>0.132</td>
<td>0.093</td>
<td>0.13</td>
<td>0.009</td>
<td>9.296</td>
<td>0.009</td>
<td>37.515</td>
</tr>
<tr>
<td>NFRS</td>
<td>69.671</td>
<td>6.975</td>
<td>206.063</td>
<td>4.36</td>
<td>0.062</td>
<td>0.037</td>
<td>2.907</td>
<td>0.465</td>
<td>2.038</td>
<td>0.013</td>
<td>31.611</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Accuracy results of algorithms with C4.5 classifiers

<table>
<thead>
<tr>
<th>Datasets</th>
<th>RMdps</th>
<th>WRMDPS</th>
<th>L - FRPS</th>
<th>FDM</th>
<th>NFRS</th>
<th>fullset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc ± Std</td>
<td>84.23</td>
<td>83.03</td>
<td>75.3</td>
<td>78.94</td>
<td>79.35</td>
<td>79.39</td>
</tr>
<tr>
<td>p-Value</td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Lose/Win/Tie</td>
<td>2/1/9</td>
<td>5/3/4</td>
<td>8/1/3</td>
<td>4/3/5</td>
<td>10/1/1</td>
<td>5/3/4</td>
</tr>
</tbody>
</table>

Table 5: Accuracy results of algorithms with NaiveBayes classifiers

<table>
<thead>
<tr>
<th>Datasets</th>
<th>RMdps</th>
<th>WRMDPS</th>
<th>L - FRPS</th>
<th>FDM</th>
<th>NFRS</th>
<th>fullset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc ± Std</td>
<td>84.23</td>
<td>83.03</td>
<td>75.3</td>
<td>78.94</td>
<td>79.35</td>
<td>79.39</td>
</tr>
<tr>
<td>p-Value</td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Lose/Win/Tie</td>
<td>2/1/9</td>
<td>5/3/4</td>
<td>8/1/3</td>
<td>4/3/5</td>
<td>10/1/1</td>
<td>5/3/4</td>
</tr>
</tbody>
</table>

Table 6: Running time results of the first step (seconds)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>alma</th>
<th>anneal</th>
<th>clean</th>
<th>credit</th>
<th>ecoli</th>
<th>glass</th>
<th>hepa</th>
<th>horse</th>
<th>ions</th>
<th>lymph</th>
<th>segm</th>
<th>AverageTime</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMdps</td>
<td>0.823</td>
<td>1.365</td>
<td>1.186</td>
<td>0.108</td>
<td>0.75</td>
<td>0.008</td>
<td>0.105</td>
<td>0.055</td>
<td>0.096</td>
<td>0.003</td>
<td>6.164</td>
<td>0.888</td>
</tr>
<tr>
<td>WRMDPS</td>
<td>0.699</td>
<td>1.384</td>
<td>1.156</td>
<td>0.12</td>
<td>0.706</td>
<td>0.013</td>
<td>0.005</td>
<td>0.054</td>
<td>0.091</td>
<td>0.003</td>
<td>6.038</td>
<td>0.864</td>
</tr>
<tr>
<td>FDM</td>
<td>0.616</td>
<td>1.415</td>
<td>1.155</td>
<td>0.115</td>
<td>0.729</td>
<td>0.013</td>
<td>0.005</td>
<td>0.053</td>
<td>0.095</td>
<td>0.003</td>
<td>6.416</td>
<td>0.894</td>
</tr>
</tbody>
</table>

1063-6706 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
In this section, we compare the proposed methods with several representative algorithms. The summary information of the experimental datasets is shown in Table 2. Colon dataset and Hepatocellular dataset are two tumor datasets. Colon dataset can be downloaded at http://www.molbio.princeton.edu/conlondata. One can find Hepatocellular carcinoma dataset in [43] (simply written as hepatocellular). AMLALL data set can be downloaded at Keng Ridge Bio-medical (KRBM) Data Set Repository. The other datasets are taken from UCI.

In this paper, we use WEKA to complete the missing values at first. The details of the hardware condition and software environment are specified as follows:

- The hardware environment: Intel(R) Core(TM) CPU 3.20GHz 8.00GB Memory.

Afterwards, we conduct our comparison experiments from two aspects: one is the comparison of efficiency (i.e. the time consumption of attribute selection), the other is the comparison of classification performance (i.e. the quality of the selected attributes). In the following, we compare the proposed algorithms RMDPS and WRMDPS with four representative algorithms: L-FRFS and FDM [33], SPS [34], NFRS [16]. L-FRFS is based on fuzzy lower approximation. It uses the fuzzy positive region to construct dependency degree, then uses dependency degree to gauge subset quality and gets the reduct. FDM is a fuzzy discernibility matrix-based feature selection algorithm. It extends the discernibility matrix to the fuzzy case, and uses individual satisfaction of each clause for a given set of attributes to find reducts. SPS is also an attribute reduction method from the viewpoint of object pair. Different from our methods, it is an algorithm based on crisp discernibility matrix generated by cut set technology, rather than fuzzy discernibility matrix.

![Figure 3: Number of remaining object pairs in i-th Cycle](image)

Table 7: Running time results of the second step (seconds)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>alma</th>
<th>anneal</th>
<th>clean</th>
<th>colon</th>
<th>credit</th>
<th>ecoli</th>
<th>glass</th>
<th>hepat</th>
<th>horse</th>
<th>ionos</th>
<th>lymph</th>
<th>segme</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMDPS</td>
<td>0.084</td>
<td>0.05</td>
<td>0.08</td>
<td>0.017</td>
<td>0.051</td>
<td>0.003</td>
<td>0.002</td>
<td>0.017</td>
<td>0.007</td>
<td>0.007</td>
<td>0.001</td>
<td>0.432</td>
<td>0.063</td>
</tr>
<tr>
<td>WRMDPS</td>
<td>0.078</td>
<td>0.049</td>
<td>0.08</td>
<td>0.014</td>
<td>0.05</td>
<td>0.003</td>
<td>0.002</td>
<td>0.016</td>
<td>0.007</td>
<td>0.007</td>
<td>0.001</td>
<td>0.536</td>
<td>0.075</td>
</tr>
<tr>
<td>SPS</td>
<td>0.124</td>
<td>0.144</td>
<td>0.139</td>
<td>0.022</td>
<td>0.178</td>
<td>0.01</td>
<td>0.005</td>
<td>0.025</td>
<td>0.022</td>
<td>0.017</td>
<td>0.004</td>
<td>0.924</td>
<td>0.1345</td>
</tr>
</tbody>
</table>

Table 8: Running time results of the third step (seconds)

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>alma</th>
<th>anneal</th>
<th>clean</th>
<th>colon</th>
<th>credit</th>
<th>ecoli</th>
<th>glass</th>
<th>hepat</th>
<th>horse</th>
<th>ionos</th>
<th>lymph</th>
<th>segme</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMDPS</td>
<td>0.054</td>
<td>0.022</td>
<td>0.006</td>
<td>0.011</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.024</td>
<td>0.009</td>
<td>0.011</td>
<td>0.001</td>
<td>0.367</td>
<td>0.048</td>
</tr>
<tr>
<td>WRMDPS</td>
<td>0.016</td>
<td>0.035</td>
<td>0.11</td>
<td>0.018</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.024</td>
<td>0.009</td>
<td>0.011</td>
<td>0.001</td>
<td>0.461</td>
<td>0.191</td>
</tr>
<tr>
<td>SPS</td>
<td>0.029</td>
<td>0.193</td>
<td>0.097</td>
<td>0.008</td>
<td>0.129</td>
<td>0.02</td>
<td>0.008</td>
<td>0.004</td>
<td>0.016</td>
<td>0.017</td>
<td>0.002</td>
<td>1.956</td>
<td>0.206</td>
</tr>
</tbody>
</table>
In essence, SPS transforms the framework of fuzzy rough sets into that of crisp rough sets. NFRRS is a fitting model feature selection algorithm for fuzzy rough sets. First, it defines the fuzzy decision of a sample using the concept of fuzzy neighborhood. Then, a parameterized fuzzy relation is introduced to characterize the fuzzy information granules. Finally, it defines the significance measure of a candidate attribute and designs a greedy forward searching strategy. As for the compared methods, original proposals of these methods have been used in our comparison.

In the experiments, we use the following similarity measure to obtain fuzzy similarity relations:

$$\mu_a(x, y) = \max\left(\min\left(\frac{f(a, y) - f(a, x) + \sigma_a}{\sigma_a}, 0\right), 1 - \frac{f(a, x) + \sigma_a - f(a, y)}{\sigma_a}\right)$$

(54)

As for the nominal (or symbolic) attributes, we use the following formula to get the similarity relations.

$$\mu_a(x, y) = \begin{cases} 0, & \text{if } f(a, y) \neq f(a, x) \\ 1, & \text{if } f(a, y) = f(a, x) \end{cases}$$

(55)

Table 3 represents the average consuming time of 10 independent running of every algorithm. It is obvious that both RMDPS and WRMDPS are more efficient than compared algorithms.

In Tables 4 and 5, we illustrate the comparison results of the classification performance with respect to the selected reducts. The results are average values of 10 folds cross validation of C4.5 and NaiveBayes. Student’s paired two-tailed t-Test is applied to evaluate the statistical significance of the difference between two averaged accuracy values: one resulted from RMDPS and the other resulted from the other algorithms. In this experimental study, we set the statistical significance to the default value 0.05. p-Val indicates the probability associated with the t-Test. The smaller the value, the more significant the difference between the two average values is. What’ more, the symbols “+” and “−” represent that the corresponding approach statistically significantly (at 0.05 level) wins and loses the competing with our RMDPS respectively. The symbol “=” represents ties.

Table 4 shows the classify performance by C4.5. Table 4 indicates that the proposed methods outperform compared algorithms in general. The proposed methods RMDPS and WRMDPS get the highest average performance on all the datasets, and they are the only methods obtain average classification accuracies over 80% except for full attribute set. Table 5 shows the classification performance by NaiveBayes classifier. From the table, we can get similar conclusion to Table 4. RMDPS and WRMDPS are still the methods which get the best average performance on all the datasets. On the whole, one may conclude that the presented methods outperform the compared algorithms and full attribute set. In the last rows of Tables 4 and 5, the statistical significance results are summarized over all compared algorithms. The results indicate that RMDPS and WRMDPS perform closely. Compared with other methods, RMDPS wins more and losses less. On the whole, proposed RMDPS and WRMDPS obtain satisfactory results.
Since SPS is also an attribute reduction approach from the viewpoint of object pairs, we need to compare the proposed methods with SPS specially.

RMDPS, WRMDPS and SPS can be broken down into three steps, shown in Fig. 1. They are similar in the first step, so the running time results are close in Table 6. RMDPS, WRMDPS and SPS differ in the second and third steps. In the second step, SPS has to convert a fuzzy discernibility matrix into a crisp discernibility matrix, which is a time-consuming process. Also from Table 7 we can see that our algorithms are always faster than SPS. As for the third step, we know that SPS has more object pairs to be evaluated from Fig. 3. On the whole, from Table 8, one can conclude that RMDPS and WRMDPS are faster than SPS. However, we should notice that SPS is faster than RMDPS and WRMDPS on few datasets. To understand this situation, we need to notice that RMDPS and WRMDPS adopt a greedy search strategy to choose the next attribute, i.e., they choose the attribute with the maximal pairNum (remainder discerned object pairs) from the remainder conditional attributes in each loop. RMDPS and WRMDPS terminate when all the discernibility pairs are covered. SPS faces the similar situation, i.e., it selects attributes gradually till all the discernibility pairs are covered. It should be noticed that SPS uses an approximate and simple search strategy by neglecting the change of remainder attributes’ discerning number in remainder un-covered pairs. In other words, SPS sorts attributes once and adds attributes one by one according to the sorted sequence. Actually, for a remainder attribute, the discerning number in remainder un-covered pairs may be changed when an attribute is selected. The order of attributes may change. According to our understanding, SPS uses such an approximate search strategy because that remaining un-covered pairs are large at each iteration.

6 Conclusion

In this paper, we first propose two concepts of the reduced maximal discernibility pairs and the minimal indiscernibility pairs. Consequently, we develop two effective algorithms based on the proposed concepts, denoted by RMDPS and WRMDPS. RMDPS and WRMDPS are attribute reduction algorithms from the viewpoint of object pair in the framework of fuzzy rough sets. They only need to deal with part of the object pairs rather than the whole object pairs from the discourse, which makes such algorithms efficient for attribute reduction. Numerical experiments are conducted and the results verify the theoretical analysis. Comparison results indicate that the proposed algorithms are effective and feasible.

In this paper, we use a normal heuristic method to choose an attribute. Actually, further study can consider other search mechanisms, such as Davis-Logemann-Loveland based strategy [44], Johnson Reducer approach [44], probabilistic search [45] and global search based on swarm intelligence [46].

The main purpose of this study is to present a method to attribute reduction issue in fuzzy rough framework from the viewpoint of object pair. We think the presented study can supply optional angle to consider attribute reduction issue, since most existing attribute selection algorithms mainly take the angle of attribute set. However, one should notice that there are also improved and optimized approaches for attribute reduction from the viewpoint of attribute set, such as efficient positive region-based approaches [47], [48], [49]. To our understanding, approaches from these two viewpoints are inspiring to each other. In the future, we plan to introduce speeding up techniques in attribute reduction from the viewpoint of attribute set into the proposed framework.

The presented study focuses on finding one reduct. In the future, we also plan to study the methods to get all the reducts of a given decision table from the viewpoint of object pair.

7 Appendix

As shown in Algorithms 1 and 2, the While circulation should run exactly \(|\text{Red}| + 1\) cycles. Suppose \(\text{maxNum}_i\) is the value of \(\text{maxNum}\) in \(i - \theta\) cycles, \(selAtt_i\) represents the selected attribute in \(i - \theta\) selection and \(\text{MDP}_i^*\) represents the the number of remaining \(\text{MDP}_{\theta-1}^*\). D. Note that \(\text{maxNum}_0 = \text{maxNum}|\text{Red}| + 1 = 0, \text{MDP}_0^*|\text{Red}| + 1 = 0\). Then we have

\[
\text{maxNum}_1 + \text{maxNum}_2 + \ldots + \text{maxNum}_{|\text{Red}|} = |\text{MDP}_C^* D(U)|
\]

\[
\text{MDP}_i^* = |\text{MDP}_C^* D(U)| - (\text{maxNum}_1 + \ldots + \text{maxNum}_{|\text{Red}| - 1}), i \geq 1
\]

We randomly select six data sets to show the changes of the values of \(\text{MDP}_i^*\) in the circulation in Fig. 3. As we can see from Fig. 3, \(\text{MDP}_i^*\) is much smaller than \(\text{MDP}_{i-1}^*\). In order to describe the changes of \(\text{MDP}_i^*\) better, we assume it follows a geometrical change, i.e. \(\text{MDP}_i^* = \text{MDP}_i^* \cdot \delta\), where \(\delta\) is a common ratio. Thus \(\text{MDP}_i^*\) can be represented as follows:

\[
\text{MDP}_i^* = \text{MDP}_1^* \cdot \delta^{i-1}
\]

Here \(\text{MDP}_1^* = \frac{|U|(|U|-1)}{2}\). Considering the value of \(\text{MDP}_0^*|\text{Red}| + 1\) is integer, thus

\[
\text{MDP}_0^*|\text{Red}| + 1 = 0 \iff 0 \leq \text{MDP}_0^*|\text{Red}| + 1 < 1
\]

\[
\iff 0 \leq \text{MDP}_1^* \cdot \delta|\text{Red}| < 1
\]

which implies

\[
\delta < \sqrt[|\text{Red}|]{\frac{1}{\text{MDP}_1^*}} \leq \sqrt{|\text{Red}|}{\frac{2}{|U|(|U|-1)}}
\]
Here, we choose \( \vartheta = \frac{\sqrt{2}}{U(|U|-1)} \) in a most conservative approach. Obviously its range is \((0, 1)\). So
\[
MDP^*_1 + MDP^*_2 + \ldots + MDP^*_r(U|U|-1) = MDP^*_1 \vartheta + MDP^*_2 \vartheta^2 + \ldots + MDP^*_r \vartheta^r(U|U|-1)
\]
which means the time complexity for steps 3 to 18 is \(O(\frac{|U|^2|C|}{2} \cdot \frac{1}{1-\vartheta})\), in which
\[
\vartheta = \frac{\sqrt{2}}{U(|U|-1)}
\]

Proposition 11. For a given decision table \( S = < U, C \cup D > \), \(|U|\) means the number of objects in \( U \), \(|C|\) is the number of attributes in \( C \), and \(|Red|\) represents the number of attributes contained in the reduct with respect to the given decision table. Then
\[
\frac{|U|^2|C|}{2} \cdot \frac{1}{1-\vartheta} \leq |U|^2|C| \quad \text{if} \quad |U|(|U|-1) \geq 2|Red|+1.
\]

Proof: It can be easily proved using Eq. 62.

REFERENCES


Jianhua Dai received the B.Sc., M.Eng., and Ph.D. degrees in computer science from Wuhan University, Wuhan, China, in 1998, 2000, and 2003, respectively. He is currently a Professor with the Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China) and the College of Information Science and Engineering, Hunan Normal University, Changsha, Hunan, China. He has published over 80 research papers in refereed journals and conferences. His current research interests include artificial intelligence, machine learning, data mining, evolutionary computation, and soft computing.

Hu Hu received the B.Sc. degree in computer science and technology from the Hefei University of Technology, Hefei, China, in 2015. He is currently pursuing the graduation degree with the College of Computer Science and Technology, Zhejiang University, Hangzhou, China. His current research interests include machine learning, data mining, and rough sets.

Debiao Huang received the B.Sc. degree in computer science and technology from Northeast University, Shenyang, China, in 2013. He is a graduate student with the College of Computer Science, Zhejiang University. His research interests include machine learning, data mining and rough sets.

Wei-Zhi Wu received the B.Sc. degree in mathematics from Zhejiang Normal University, Jinhua, China, in 1986, the M.Sc. degree in mathematics from East China Normal University, Shanghai, China, in 1992, and the Ph.D. degree in applied mathematics from Xi’an Jiaotong University, Xi’an, China, in 2002. He is currently a Professor with the School of Mathematics, Physics, and Information Science, Zhejiang Ocean University, Zhejiang, China. His current research interests include approximate reasoning, rough sets, random sets, formal concept analysis, and granular computing.

Yuhua Qian received the M.S. degree and the Ph.D. degree in Computers with applications at Shanxi University in 2005 and 2011, respectively. He is currently a Professor with the School of Computer and Information Technology, Shanxi University, Taiyuan, China. His current research interests include pattern recognition, feature selection, rough set theory, granular computing, and artificial intelligence.

Jianhua Dai received the B.Sc., M.Eng., and Ph.D. degrees in computer science from Wuhan University, Wuhan, China, in 1998, 2000, and 2003, respectively. He is currently a Professor with the Key Laboratory of High Performance Computing and Stochastic Information Processing (Ministry of Education of China) and the College of Information Science and Engineering, Hunan Normal University, Changsha, Hunan, China. He has published over 80 research papers in refereed journals and conferences. His current research interests include artificial intelligence, machine learning, data mining, evolutionary computation, and soft computing.

Hu Hu received the B.Sc. degree in computer science and technology from the Hefei University of Technology, Hefei, China, in 2015. He is currently pursuing the graduation degree with the College of Computer Science and Technology, Zhejiang University, Hangzhou, China. His current research interests include machine learning, data mining, and rough sets.

Debiao Huang received the B.Sc. degree in computer science and technology from Northeast University, Shenyang, China, in 2013. He is a graduate student with the College of Computer Science, Zhejiang University. His research interests include machine learning, data mining and rough sets.