Maximal-Discernibility-Pair-Based Approach to Attribute Reduction in Fuzzy Rough Sets

Jianhua Dai, Hu Hu, Wei-Zhi Wu, Yuhua Qian, and 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, the fuzzy rough set model is proposed, which is an extended model of rough sets and is able to deal with imprecision and uncertainty in both symbolic and numerical attributes. The existing attribute selection algorithms based on the fuzzy rough set model mainly take the angle of "attribute set," which means they define the object function representing the predictive ability for an attribute subset with regard to the domain of discourse, rather than following the view of an "object pair." Algorithms from the viewpoint of the object pair can ignore the object pairs that are already discerned by the selected attribute 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 the object pair in the framework of the fuzzy rough set model. Then, we develop two attribute selection algorithms, named as reduced maximal discernibility pairs selection and weighted reduced maximal discernibility pair selection, based on the reduced maximal discernibility pairs. Experiment results show that the proposed algorithms are effective and efficient in attribute selection.**

***Index Terms*—Attribute reduction, fuzzy discernibility matrix, fuzzy rough sets, maximal discernibility pairs.**

I. Introduction

THE rough set theory, introduced by Pawlak [1], is a useful mathematical approach to deal with vague and uncertain information, has attracted many researchers' attention, and has been proven to be successful in solving a variety of problems [1]–[6]. However, just as [7] mentioned, a rough set model can only deal with a symbolic value, while values of attributes might be either symbolic or real valued in many real datasets. To overcome this problem, several extended models have been proposed [8], and the 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 [3], [11]–[16], which should be taken as a necessary preprocessing step to find a suitable subset of attributes.

In the fuzzy rough set model, we use the fuzzy similarity relation to replace the equivalence relation in crisp rough set theory to measure the indiscernibility between two objects. In most of the 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 research studies on a fuzzy rough set contain at least two topics: the construction of the approximations of the fuzzy rough set model and the applications of the fuzzy rough set model. On one hand, since the 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, the fuzzy rough set model has been successfully applied in many applications [19], such as classification [20], clustering [21], and rule extraction [22], especially attribute reduction [3], [23]–[27].

Attribute reduction of fuzzy rough set theory has been a popular topic in recent years. Shen and Jensen [28] generalized the dependence function defined in the classical rough set based on the positive region into the fuzzy case and presented a fuzzy rough attribute selection algorithm based on such dependence. In [29], Bhatt and Gopal 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 datasets. In [32] and [33], a fuzzy discernibility matrix was proposed, and the algorithm based on dependence [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

Manuscript received January 30, 2017; revised June 9, 2017 and September 14, 2017; accepted October 11, 2017. Date of publication October 30, 2017; date of current version August 2, 2018. This work was supported in part by the National Natural Science Foundation of China (Nos. 61473259, 61573321, 41631179, 61070074, and 60703038), the Zhejiang Provincial Natural Science Foundation of China (Nos. LY14F020029 and LY18F030017), and the National Science and Technology Support Program of China (2015BAK26B00 and 2015BAK26B02). *(Corresponding author: Jianhua Dai.)*

J. Dai is 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 410081, China, and also with the School of Computer Science and Technology, Tianjin University, Tianjin 300350, China (e-mail: david.joshua@qq.com).

H. Hu and D. Huang are with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: 2420967879@qq.com; 1169196994@qq.com).

W.-Z. Wu is with the School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhejiang 316022, China (e-mail: wzwu@zjou.edu.cn).

Y. Qian is with the School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China (e-mail: jinchengqyh@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2017.2768044

concept of fuzzy similarity measure and a model for evaluating feature dependence were presented. In [36], an accelerator, called forward approximation, was constructed by combining sample reduction and dimensionality reduction together.

In summary, the existing attribute reduction methods based on the fuzzy rough set model mainly adopt the angle of an "attribute set," which means they use an object function to represent the predictive ability of the attribute subset with regard to the domain of discourse, rather than the view of an "object pair." Comparatively speaking, algorithms based on the "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 the crisp discernibility matrix generated by cut set technology, rather than the 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 the "object pair" is directly introduced into the framework of fuzzy rough sets. Then, we develop two attribute selection algorithms, named as reduced maximal discernibility pair selection (RMDPS) and weighted reduced maximal discernibility pair selection (WRMDPS), 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 II. The related definitions and concepts of reduced maximal discernibility pairs are proposed in Section III. The proposed attribute significance measure and attribute reduction algorithms based on reduced maximal discernibility pairs are introduced in Section IV. Experiments are conducted in Section V, and Section VI concludes this paper.

II. 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.

A. Rough Sets

A decision table is defined as $S =< U, A, V, f >$, where $U = \{x_1, x_2, ..., x_n\}$ is a finite nonempty set of objects; $A = C \cup D$ is a finite nonempty set of attributes, where $C = \{c_1, c_2, ..., c_m\}$ 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., $V = \cup_{a \in A} V_a$, where V_a is the value set of attribute a, called the value domain for attribute a; and $f : U \times A \to V$ is an information function, which maps an object in U to exactly one value from domains of attribute such as $\forall a \in A, x \in U$, and $f(a, x) \in V_a$, where $f(a, x)$ represents the value of object x on attribute a.

Given a decision table $S =< U, A, V, f >$, for any subset of attributes $B \subseteq A$, the indiscernibility relation generated by B on U is defined by

$$\text{IND}(B) = \{(x, y) \in U^2 | f(b, x) = f(b, y), \forall b \in B\} \quad (1)$$

It is clear that $\text{IND}(B)$ is an equivalence relation, which is reflexive, symmetric, and transitive. It determines a partition of U, denoted by $U/\text{IND}(B)$ or simply U/B; an equivalence class of $\text{IND}(B)$ 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

$$\underline{R}_B X = \{x | [x]_B \subseteq X\} \quad (2)$$

$$\overline{R}_B X = \{x | [x]_B \cap X \neq \emptyset\}. \quad (3)$$

$\underline{R}_B X$ and $\overline{R}_B X$ are two key concepts in rough set theory. Suppose that 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:

$$\text{POS}_P\{Q\} = \underset{X \in U/Q}{\cup} \underline{R}_P X \quad (4)$$

$$\text{NEG}_P\{Q\} = U - \underset{X \in U/Q}{\cup} \overline{R}_P X \quad (5)$$

$$\text{BND}_P\{Q\} = \underset{X \in U/Q}{\cup} \overline{R}_P X - \underset{X \in U/Q}{\cup} \underline{R}_P X. \quad (6)$$

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 $S =< U, C \cup D >$, a subset $B \subseteq C$ is called a relative reduct of C if B is independent in S, and $\text{POS}_B(D) = \text{POS}_C(D)$. The set of all indispensable attributes in C is called the core and denoted by $\text{Core}_U(C \cup D)$. The set of all reducts is denoted as $\text{Red}_U(C \cup D)$, and we have $\text{Core}_U(C \cup D) = \cap \text{Red}_U(C \cup D)$.

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

$$\gamma_B(D) = \frac{|\text{POS}_B(D)|}{|U|} \quad (7)$$

$\gamma_B(D) = 1$ 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 on 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 γ. In other words, a subset of conditional attributes can be regarded as a reduct only if it satisfies

$$1)\ \gamma_B(D) = \gamma_C(D) \quad (8)$$

$$2)\ \forall B' \subset B, \gamma_{B'}(D) < \gamma_B(D). \quad (9)$$

There might exist many reducts for a dataset. 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

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

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 $\text{Core}_U(C \cup D)$ iff $\exists x, y \in U$ satisfying $M(x, y) = \{c_i\}$. Thus, in the discernibility matrix, the core is defined by

$$\text{Core}_U(C \cup D) = \cup\{M(x,y) | x, y \in U \text{ and } |M(x,y)| = 1\} \tag{11}$$

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:

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

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

$$1)\ \forall x, y \in U, M(x,y) \neq \emptyset \rightarrow B \cap M(x,y) \neq \emptyset \tag{13}$$

$$2)\ \forall B' \subset B, \exists x, y \in U, M(x,y) \neq \emptyset \rightarrow B' \cap M(x,y) = \emptyset. \tag{14}$$

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

B. Fuzzy Rough Sets

The traditional rough set theory can only process symbolic-valued attributes. However, most of the real datasets 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 the fuzzy rough set model [38], [39], we can handle the real-valued attributes or the hybrid attributes directly.

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

1) Reflexivity: $\forall x \in U, S(x, x) = 1$.
2) Symmetry: $\forall x, y \in U, S(x, y) = S(y, x)$.
3) 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 $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ and holds the following conditions.

1) Commutativity: $T(a, b) = T(b, a)$.
2) Monotonicity: $T(a, b) \leq T(c, d)$, if $a \leq c$ and $b \leq d$.
3) Associativity: $T(a, T(b, c)) = T(T(a, b), c)$.
4) 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 and Kerre proposed the lower and upper approximations by means of the T-transitive fuzzy similarity relation:

$$\mu_{\underline{R}_P} X(x) = \inf_{y \in U} I(\mu_{R_P}(x, y), \mu_X(y)) \tag{15}$$

$$\mu_{\overline{R}_P} X(x) = \sup_{y \in U} T(\mu_{R_P}(x, y), \mu_X(y)) \tag{16}$$

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

$$\mu_{R_P}(x, y) = \min_{a \in P}\{\mu_{R_a}(x, y)\} \tag{17}$$

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 the 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:

$$\mu_{\text{POS}_Q}(x) = \sup_{X \in U/Q} \mu_{\underline{R}_P} X(x). \tag{18}$$

The fuzzy rough dependence is defined by

$$\gamma'_P(Q) = \frac{|\text{POS}_P(Q)|}{|U|} = \frac{\sum_{x \in U} \mu_{\text{POS}_Q}(x)}{|U|}. \tag{19}$$

With fuzzy dependence, we are able to measure the ability of a given subset of attributes for preserving the dependence degree of the entire attributes. By means of comparing fuzzy dependence, 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

$$1)\ \gamma'_B(D) = \gamma'_C(D) \tag{20}$$

$$2)\ \forall B' \subset B, \gamma_{B'}(D) < \gamma_B(D). \tag{21}$$

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:

$$M'(x,y) = \{a_{N(\mu_{R_a}(x,y))} | a \in C\} \ \ \forall x, y \in U \tag{22}$$

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 $\{a_{0.42}, b_{0.548}, c_{1.0}\}$. The $\text{Core}'_U(C)$ in the fuzzy discernibility matrix is defined by

$$\text{Core}'_U(C) = \{a | \exists M'(x,y), \mu_{R_a}(x,y) > 0, \forall c \in C - \{a\}, \mu_{R_c}(x,y) = 0\}. \tag{23}$$

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

$$f'_D(c_1, c_2, ..., c_m) = \wedge\{\vee M'(x,y) \leftarrow N(\mu_{R_D}(x,y)) | \forall x, y \in U\}. \tag{24}$$

Here, $\leftarrow$ represents the fuzzy implication and D means the decision attributes. It is noteworthy that, the same as the discernibility function, the satisfaction of the clause is largely affected by the value of decision attributes. In this paper, we concentrate on the datasets that only contain single symbolic decision attribute. Thus, $\mu_{R_D}(x,y)$ takes values in $\{0,1\}$, i.e., $\mu_{R_D}(x,y) = 1$ means objects x and y have the same decision value; otherwise, objects x and y have different decision values.

In order to find reducts, the degree of satisfaction of a clause $M'(x,y)$ for a subset of attributes B with respect to decision attribute D is proposed:

$$\text{SAT}_{B,D}(M'(x,y)) = \max_{a \in B}\{N(\mu_{R_a}(x,y))\} \leftarrow N(\mu_{R_D}(x,y)). \tag{25}$$

Thus, the total satisfiability of all clauses for B is defined as

$$\text{SAT}(B) = \frac{\sum_{x,y \in U, x \neq y} \text{SAT}_{B,D}(M'(x,y))}{\sum_{x,y \in U, x \neq y} \text{SAT}_{C,D}(M'(x,y))}. \tag{26}$$

Once $\text{SAT}(B) = 1$, then the subset B is a fuzzy rough reduct. In other words, a subset of conditional attributes B can be regarded as a reduct only if it satisfies the following:

$$1)\ \text{SAT}(B) = \text{SAT}(C) = 1 \tag{27}$$

$$2)\ \forall B' \subset B, \text{SAT}(B') < \text{SAT}(B). \tag{28}$$

III. Maximum Discernibility Pairs

In a crisp discernibility matrix, as (13) and (14) indicate, a reduct B is a minimum attribute subset that overlaps each entry in the crisp discernibility matrix with at least one attribute. If we take each entry in the crisp discernibility matrix as a basic unit and neglect the matrix that contains those entries, then new algorithms (such as the algorithm in [34]) can be proposed. Since each entry can be represented by an object pair (x,y), we can also call such algorithms based on the "object pair." One important property in attribute reduction based on the 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 a pair can be discerned by any attribute in it.

We first introduce two important concepts in fuzzy rough sets:

$$\text{sim}_a(x,y) = \mu_a(x,y), \qquad a \in C \tag{29}$$

$$\text{dis}_a(x,y) = N(\mu_a(x,y)), \qquad a \in C. \tag{30}$$

$\text{sim}_a(x,y)$ represents the fuzzy indiscernibility between objects x and y; on the contrary, $\text{dis}_a(x,y)$ 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:

$$\text{sim}_a(x,y) = \text{sim}_a(y,x). \tag{31}$$

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

$$\text{SAT}_{B,D}(M'(x,y)) = \max_{a \in B}\{1 - \mu_{R_a}(x,y)\} \leftarrow N(\mu_{R_D}(x,y)) \tag{32}$$

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:

$$\text{SAT}_{B,D}(M'(x,y)) = \min_{a \in B}\{\mu_{R_a}(x,y)\} \leftarrow N(\mu_{R_D}(x,y)) \tag{33}$$

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 are defined as follows:

$$\text{MSA}_C D(x,y) = \{a | \text{sim}_a(x,y) = \min_{a \in C} \text{sim}_a(x,y), a \in C\} \leftarrow\!\!\!\rho\ N(\mu_{R_D}(x,y)) \tag{34}$$

where $\leftarrow\!\!\!\rho$ is a symbol representing a switch used to determine whether its left side should be calculated, i.e., $a \leftarrow\!\!\!\rho\ b$ 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, (34) is equal to the following equation:

$$\text{MSA}_C D(x,y) = \begin{cases} \{a | \text{sim}_a(x,y) = \min_{c \in C} \text{sim}_c(x,y), \\ a \in C\}, & \text{if } \mu_{R_D}(x,y) = 0 \\ \emptyset, & \text{if } \mu_{R_D}(x,y) = 1. \end{cases} \tag{35}$$

The maximal fuzzy discernibility attributes corresponding to pair (x, y) in the fuzzy discernibility matrix is defined as

$$\text{MDA}_C D(x,y) = \{a | \text{dis}_a(x,y) = \max_{a \in C} \text{dis}_c(x,y), a \in C\} \\ \hookleftarrow N(\mu_{R_D}(x,y)). \tag{36}$$

In this paper, we use $|\text{MSA}_C(x,y)|$ and $|\text{MDA}_C(x,y)|$ to represent the number of conditional attributes contained in the minimal fuzzy similarity attributes and the maximal fuzzy discernibility attributes, respectively.

Equation (36) is equal to the following equation:

$$\text{MDA}_C D(x,y) = \begin{cases} \{a | \text{dis}_a(x,y) = \max_{c \in C} \text{dis}_c(x,y), \\ a \in C\}, & \text{if } \mu_{R_D}(x,y) = 0 \\ \emptyset, & \text{if } \mu_{R_D}(x,y) = 1. \end{cases} \tag{37}$$

The above definition can successfully degrade to the crisp form, in which $\text{MSA}_C D(x,y)$ and $\text{MDA}_C D(x,y)$ 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 $S =< U, C \cup D >$, for any pair (x,y) in the corresponding fuzzy discernibility matrix, we have

$$\text{MDA}_C D(x,y) = \text{MDA}_C D(y,x) \tag{38}$$

$$\text{MSA}_C D(x,y) = \text{MSA}_C D(y,x) \tag{39}$$

$$\text{MDA}_C D(x,y) = \text{MSA}_C D(x,y). \tag{40}$$

Proof: Since the fuzzy discernibility matrix is symmetric, according to Definition 1, (38) and (39) can be easily obtained. The result in (40) follows directly by using (29), (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 $S =< U, C \cup D >$, the minimal fuzzy similarity pairs with respect to attribute $a \in C$ can be defined by

$$\text{MSP}_a D(U) = \{(x,y) | a \in \text{MSA}_C D(x,y), (x,y) \in U \times U\} \tag{41}$$

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

$$\text{MDP}_a D(U) = \{(x,y) | a \in \text{MDA}_C D(x,y), (x,y) \in U \times U\}. \tag{42}$$

In the following, we use $|\text{MSP}_a D(U)|$ and $|\text{MDP}_a D(U)|$ to represent the sizes of the minimal fuzzy similarity pairs and the maximal fuzzy discernibility pairs, respectively.

Proposition 2: For a given decision table $S =< U, C \cup D >$, for any attribute $a \in C$, we have

$$\text{MSP}_a D(U) = \text{MDP}_a D(U). \tag{43}$$

Proof: From Proposition 1, we know that $\text{MDA}_C D(x,y) = \text{MSA}_C D(x,y)$. According to Definition 2, it is easy to get $\text{MSP}_a D(U) = \text{MDP}_a D(U)$. ■

TABLE I
DECISION TABLE

U	c_1	c_2	c_3	d
x_1	0.2	0.3	− 0.2	true
x_2	0.3	0.2	0.1	false
x_3	− 0.1	0	0	true
x_4	− 0.2	− 0.1	0	false

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

$$\text{sim}_{c_1} = \begin{pmatrix} 1.000 & 0.580 & 0.000 & 0.000 \\ 0.580 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.580 \\ 0.000 & 0.000 & 0.580 & 1.000 \end{pmatrix}$$

$$\text{sim}_{c_2} = \begin{pmatrix} 1.000 & 0.452 & 0.000 & 0.000 \\ 0.452 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.452 \\ 0.000 & 0.000 & 0.452 & 1.000 \end{pmatrix}$$

$$\text{sim}_{c_3} = \begin{pmatrix} 1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.205 & 0.205 \\ 0.000 & 0.205 & 1.000 & 1.000 \\ 0.000 & 0.205 & 1.000 & 1.000 \end{pmatrix}.$$

Thus

$$\text{MDP}_{c_1} D(U) = \{(x_1,x_4),(x_4,x_1),(x_2,x_3),(x_3,x_2)\}$$

$$\text{MDP}_{c_2} D(U) = \{(x_1,x_4),(x_4,x_1),(x_2,x_3),(x_3,x_2), \\ (x_3,x_4),(x_4,x_3)\}$$

$$\text{MDP}_{c_3} D(U) = \{(x_1,x_2),(x_2,x_1),(x_1,x_4),(x_4,x_1)\}$$

$$\text{MSP}_{c_1} D(U) = \{(x_1,x_4),(x_4,x_1),(x_2,x_3),(x_3,x_2)\}$$

$$\text{MSP}_{c_2} D(U) = \{(x_1,x_4),(x_4,x_1),(x_2,x_3),(x_3,x_2), \\ (x_3,x_4),(x_4,x_3)\}$$

$$\text{MSP}_{c_3} D(U) = \{(x_1,x_2),(x_2,x_1),(x_1,x_4),(x_4,x_1)\}.$$

As (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 $S =< U, C \cup D >$, the maximal fuzzy discernibility pairs with respect to attribute subset $B \subseteq C$ can be defined in the following way:

$$\text{MDP}_B D(U) = \bigcup_{\forall a \in B} \text{MDP}_a D(U). \tag{44}$$

Here, $\text{MDP}_B D(U)$ is a set that contains all the maximal discernibility pairs with respect to the attribute in B.

Proposition 3: For a given decision table $S =< U, C \cup D >$, for any attribute subset $B \subseteq C$, we have

$$\text{MDP}_B D(U) \subseteq \text{MDP}_C D(U), \forall B \subseteq C \tag{45}$$

Proof: From Definition 3, it is easy to know if $B \subseteq C$, $\text{MDP}_B D(U) \subseteq \text{MDP}_C D(U)$. ■

Proposition 4: For a given decision table $S =< U, C \cup D >$, $\forall B' \subseteq B \subseteq C$, $\text{MDP}_{B'} D(U) \subseteq \text{MDP}_B D(U)$.

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

According to Proposition 4, we know that $|\text{MDP}_B D(U)|$ satisfies monotonicity with respect to attribute subset B.

Proposition 5: For a given decision table $S =< U, C \cup D >$, $\forall B' \subseteq B$, $\forall a \in C - B$, we have

$$|\text{MDP}_{B'} D(U)| \leq |\text{MDP}_B D(U)| \tag{46}$$

$$|\text{MDP}_B D(U)| \leq |\text{MDP}_{B \cup \{a\}} D(U)|. \tag{47}$$

Proposition 6: For a given decision table $S= < U, C \cup D >$, $\text{MDP}_B D(U)$=$\text{MDP}_C D(U)$ iff $|\text{MDP}_B D(U)|$=$|\text{MDP}_C D(U)|$.

Proof: According to (45) in Proposition 3, $\text{MDP}_B D(U) = \text{MDP}_C D(U) \Leftrightarrow |\text{MDP}_B D(U)| = |\text{MDP}_C D(U)|$. Thus, the proposition holds. ■

Proposition 7: For a given decision table $S =< U, C \cup D >$, B is a reduct of the given decision table if it satisfies the following.

1) $\text{MDP}_B D(U) = \text{MDP}_C D(U)$.
2) $\forall B' \subset B, \text{MDP}_{B'} D(U) \subset \text{MDP}_B D(U)$.

Proof: According to review of Section II-B, B is a reduct if it satisfies (27) and (28).

Once we use $N(x) = 1 - x$ and operator max to specify it, by (32), we have

$$\begin{aligned}
&\text{SAT}(B) \\
&= \frac{\sum_{x,y \in U, x \neq y} \text{SAT}_{B,D}(M'(x,y))}{\sum_{x,y \in U, x \neq y} \text{SAT}_{C,D}(M'(x,y))} \\
&= \frac{\sum_{x,y \in U, x \neq y} \max\limits_{a \in B} \{1 - \mu_{R_a}(x,y)\} \looparrowright N(\mu_{R_D}(x,y))}{\sum_{x,y \in U, x \neq y} \max\limits_{a \in C} \{1 - \mu_{R_a}(x,y)\} \looparrowright N(\mu_{R_D}(x,y))}
\end{aligned} \tag{48}$$

which means

$$\begin{aligned}
&\text{SAT}(B) = \text{SAT}(C) \\
&\Leftrightarrow \sum_{x,y \in U, x \neq y} \max_{a \in B} \{1 - \mu_{R_a}(x,y)\} \looparrowright N(\mu_{R_D}(x,y)) \\
&= \sum_{x,y \in U, x \neq y} \max_{a \in C} \{1 - \mu_{R_a}(x,y)\} \looparrowright N(\mu_{R_D}(x,y)) \\
&\Leftrightarrow \sum_{x,y \in U, x \neq y} \text{dis}_a(x,y),\ \forall a \in \text{MDA}_C D(x,y) \\
&= \sum_{x,y \in U, x \neq y} \text{dis}_a(x,y),\ \forall a \in \text{MDA}_B D(x,y) \\
&\Leftrightarrow \text{MDP}_B D(U) = \text{MDP}_C D(U).
\end{aligned} \tag{49}$$

$\text{SAT}(B') < \text{SAT}(B) \Leftrightarrow \text{MDP}_{B'} D(U) \subset \text{MDP}_B D(U)$ can be obtained by the similar method.

In summary, the proposition holds. ■

Definition 4: For a given decision table $S =< U, C \cup D >$, the reduced maximal discernibility pairs with respect to attribute $a \in C$ can be defined in the following way:

$$\begin{aligned}
\text{MDP}'_a D(U) = \{&(x_i, x_j) | a \in \text{MDA}_C D(x_i, x_j), \\
&\forall x_i, x_j \in U, i < j\}
\end{aligned} \tag{50}$$

$$\begin{aligned}
\text{MDP}^*_a D(U) = \{&(x_i, x_j) | a \in \text{MDA}_C D(x_i, x_j), \\
&\forall x_i, x_j \in U, i > j\}.
\end{aligned} \tag{51}$$

Definition 5: For a given decision table $S =< U, C \cup D >$, the reduced maximal fuzzy discernibility pairs with respect to attribute subset $B \subseteq C$ can be defined in the following way:

$$\text{MDP}'_B D(U) = \bigcup_{a \in B} \text{MDP}'_a D(U) \tag{52}$$

$$\text{MDP}^*_B D(U) = \bigcup_{a \in B} \text{MDP}^*_a D(U). \tag{53}$$

According to Definition 4, we know that $\text{MDP}'_B D(U)$ or $\text{MDP}^*_B D(U)$ are exactly the half part of $\text{MDP}_B D(U)$. Thus, our attention will be only focused on the $\text{MDP}'_B D(U)$ in the following for the purpose of simplicity.

Proposition 8: For a given decision table $S =< U, C \cup D >$, $\forall B' \subseteq B \subseteq C$, $\text{MDP}'_{B'} D(U) \subseteq \text{MDP}'_B D(U)$.

Proof: From Definition 5, we know that $\text{MDP}'_B D(U)$ is the union of $\text{MDP}'_a D(U), a \in B$. Hence, it is easy to get $\text{MDP}'_{B'} D(U) \subseteq \text{MDP}'_B D(U)$ if $\forall B' \subseteq B$. ■

Proposition 9: For a given decision table $S= < U, C \cup D >$, $\text{MDP}'_B D(U)$=$\text{MDP}'_C D(U)$ iff $|\text{MDP}'_B D(U)|$=$|\text{MDP}'_C D(U)|$.

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

Proposition 10: For a given decision table $S =< U, C \cup D >$, B is a reduct of the given decision table iff

1) $\text{MDP}'_B D(U) = \text{MDP}'_C D(U)$
2) $\forall B' \subset B, \text{MDP}'_{B'} D(U) \subset \text{MDP}'_C D(U)$.

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

Example 2: Let us consider the dataset in Table I again. Then

$\text{MDP}_{c_1,c_2} D(U) = \{(x_1,x_4), (x_4,x_1), (x_2,x_3), (x_3,x_2), (x_3,x_4), (x_4,x_3)\}$
$\text{MDP}_{c_2,c_3} D(U) = \{(x_1,x_4), (x_4,x_1), (x_2,x_3), (x_3,x_2), (x_3,x_4), (x_4,x_3), (x_1,x_2), (x_2,x_1)\}$
$\text{MDP}_{c_1,c_3} D(U) = \{(x_1,x_4), (x_4,x_1), (x_2,x_3), (x_3,x_2), (x_1,x_2), (x_2,x_1)\}$
$\text{MDP}_{c_1,c_2,c_3} D(U) = \{(x_1,x_4), (x_4,x_1), (x_2,x_3), (x_3,x_2), (x_3,x_4), (x_4,x_3), (x_1,x_2), (x_2,x_1)\}$
$\text{MDP}^*_{c_1,c_2} D(U) = \{(x_4,x_1), (x_3,x_2), (x_4,x_3)\}$
$\text{MDP}^*_{c_2,c_3} D(U) = \{(x_4,x_1), (x_3,x_2), (x_4,x_3), (x_2,x_1)\}$
$\text{MDP}^*_{c_1,c_3} D(U) = \{(x_4,x_1), (x_3,x_2), (x_2,x_1)\}$
$\text{MDP}^*_{c_1,c_2,c_3} D(U) = \{(x_4,x_1), (x_3,x_2), (x_4,x_3), (x_2,x_1)\}$
$\text{MDP}'_{c_1,c_2} D(U) = \{(x_1,x_4), (x_2,x_3), (x_3,x_4)\}$
$\text{MDP}'_{c_2,c_3} D(U) = \{(x_1,x_4), (x_2,x_3), (x_3,x_4), (x_1,x_2)\}$
$\text{MDP}'_{c_1,c_3} D(U) = \{(x_1,x_4), (x_2,x_3), (x_1,x_2)\}$
$\text{MDP}'_{c_1,c_2,c_3} D(U) = \{(x_1,x_4), (x_2,x_3), (x_3,x_4), (x_1,x_2)\}$.

Characteristics of MDP and MDP′ indicate that we can evaluate the attribute subset from the viewpoint of the "object pair."

Algorithm 1: RMDPS.

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

1: Red=$\emptyset$; maxNum=0; pairNum[i]=0, $1 \le i \le |C|$.
2: Compute $\text{MDP}'_C D(U)$;
3: **while** true **do**
4: pairNum[i]=0, $1 \le i \le |C|$;
5: **for all** $(x_i, x_j) \in \text{MDP}'_C D(U)$ **do**
6: **for all** $c_k \in \text{MDA}_C D(x_i, x_j)$ **do**
7: pairNum[k]++;
8: **end for**
9: **end for**
10: maxNum=0;
11: **for all** $c_k \in C, 1 \le k \le |C|$ **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: $\text{MDP}'_C D(U) = \text{MDP}'_C D(U) - \text{MDP}'_{c_{\text{selAtt}}} D(U)$;
20: **else**
21: BREAK;
22: **end if**
23: **end while**

Algorithm 2: WRMDPS.

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

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

As for attribute reduction, methods from the angle of the 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.

IV. Algorithms Based on The Maximal Discernibility Pair Selection

According to Propositions 8–10, we find that the reduced maximal discernibility pairs are suitable for attribute selection. In this section, we propose two algorithms, called RMDPS and WRMDPS, in the framework of the reduced maximal discernibility pairs.

A. 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 the 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., selAtt, then we delete any object pair (x_i, x_j) that 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 steps 3 to 17 is $\text{O}(\frac{|U|^2|C|}{2} * \frac{1}{1-\partial})$, in which $\partial = \sqrt[|\text{Red}|]{\frac{2}{|U|(|U|-1)}}$. The detailed analysis can be found in the Appendix. In summary, the time complexity of RMDPS is $\max(O(|U|^2|C|), O(\frac{|U|^2|C|}{2} * \frac{1}{1-\partial}))$. According to Proposition 11 in the Appendix, $O(\frac{|U|^2|C|}{2} * \frac{1}{1-\partial})$ is smaller compared with $O(|U|^2|C|)$ when |red| is small. In order to make our algorithm easier to understand, we also present it in a flowchart, shown in Fig. 2.

B. 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 $|\text{MDP}_C(x_i, x_j)|$ together to measure the importance of each attribute. In other words, the importance of each attribute c_i is measured by the value of $|\text{MDP}'_{c_i} D(U)|$ and the value of $|\text{MDP}_C(x_i, x_j)|$ in which $(x_i, x_j) \in \text{MDP}'_{c_i} D(U)$.

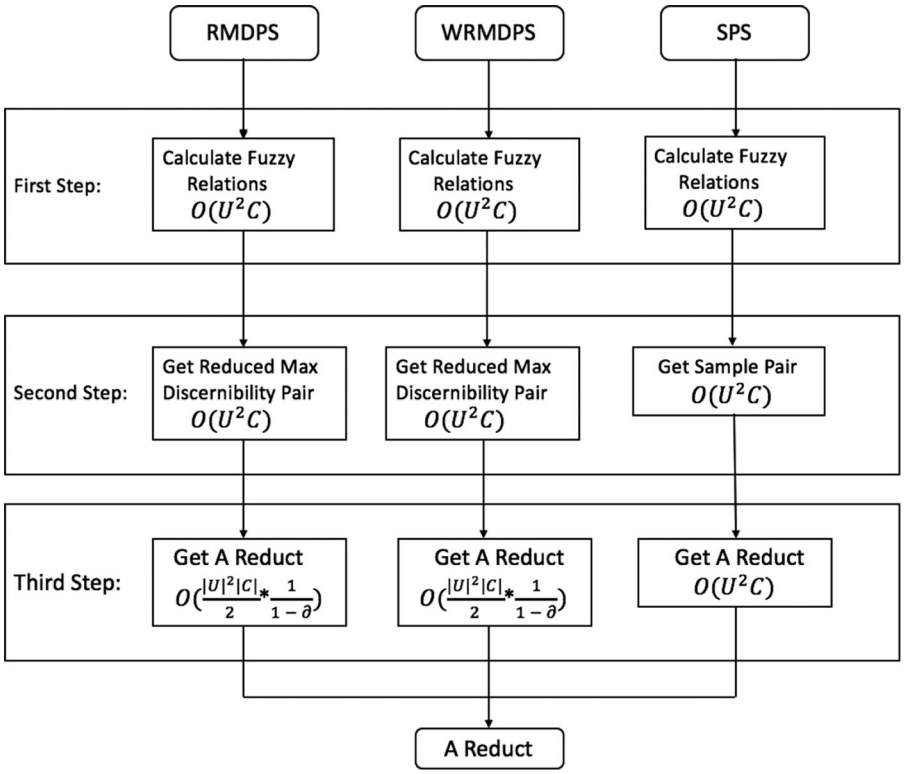

Fig. 1. Compare Structures of RMDPS, WRMDPS, and SPS.

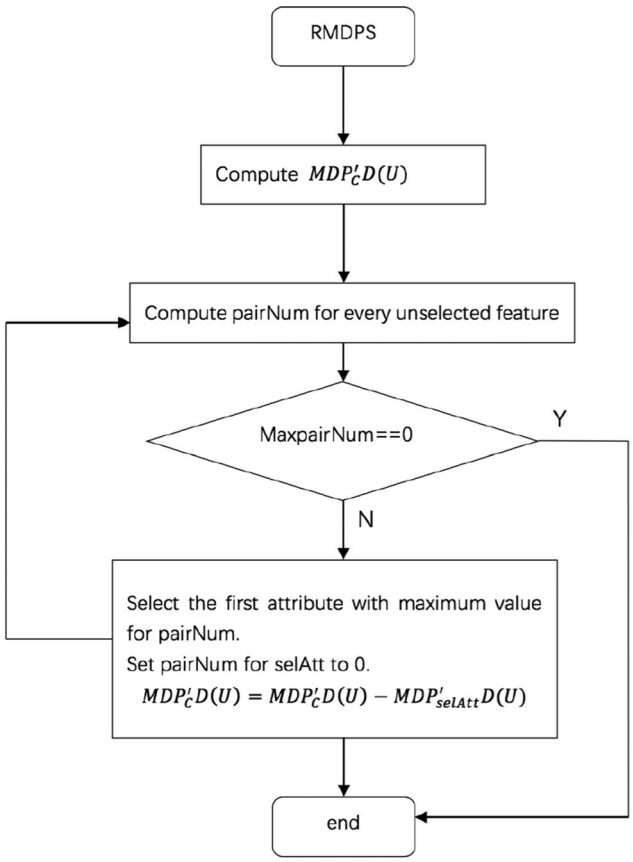

Fig. 2. Flowchart of RMDPS.

Its time complexity is $\max(O(|U|^2|C|), O(\frac{|U|^2|C|}{2} * \frac{1}{1-\partial}))$, and the detailed analysis is similar to the time complexity in Section IV-A.

V. Experiment Study

In this section, we compare the proposed methods with several representative algorithms. The summary information of the experimental datasets is shown in Table II. The Colon dataset and the Hepatocellular dataset are two tumor datasets. The Colon dataset can be downloaded at http://www.molbio.princeton.edu/conlondata. One can find the Hepatocellular carcinoma dataset in [43] (simply written as hepatocellular). The AML-ALL dataset can be downloaded at Keng Ridge Bio-medical (KRBM) Data Set Repository.[1] The other datasets are taken from UCI.[2]

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.

1) The hardware environment: Intel(R) Core(TM) CPU 3.20-GHz 8.00-GB Memory.
2) The software environment: Eclipse 3.12.1.v20160907-1200, java8.

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

[1] http://datam.i2r.a-tar.edu.sg/datasets/
[2] http://www.ics.uci.edu/mlearn/MLRepository.html

TABLE II
DETAILED INFORMATION OF THE DATASETS

Index	Dataset	Abbreviation	Objects	Conditional Attributes	Data type	Decision classes
1	Glass Identification	glass	214	9	Numeric	7
2	Horse Colic	horse	368	22	Nominal, Numeric	2
3	Ecoli	ecoli	336	7	Numeric	8
4	clean1	clean	476	166	Numeric	2
5	creditg	credit	1000	20	Nominal, Numeric	2
6	Lymphography	lymph	148	18	Numeric, Nominal	4
7	ColonAll	colon	62	2000	Numeric	2
8	Hepatocellular	hepa	33	7192	Numeric	2
9	ionosphere	ionos	351	34	Numeric	2
10	AMLALL	alma	72	7129	Numeric	2
11	segment	segm	2310	19	Numeric	7
12	anneal	anneal	798	38	Numeric, Nominal	6

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], and NFRS [16]. L-FRFS is based on fuzzy lower approximation. It uses the fuzzy positive region to construct dependence degree and then uses dependence 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 the object pair. Different from our methods, it is an algorithm based on the crisp discernibility matrix generated by cut set technology, rather than fuzzy discernibility matrix. In essence, SPS transforms the framework of fuzzy rough sets into that of crisp rough sets. NFRS 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:

$$\mu_a(x,y) = \max\left(\min\left(\frac{f(a,y)-f(a,x)+\sigma_a}{\sigma_a}, \frac{f(a,x)+\sigma_a-f(a,y)}{\sigma_a}\right), 0\right) \tag{54}$$

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

$$\mu_a(x,y) = \begin{cases} 0, \text{ if } f(a,y) \neq f(a,x) \\ 1, \text{ if } f(a,y) = f(a,x). \end{cases} \tag{55}$$

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

In Tables IV and V, we illustrate the comparison results of the classification performance with respect to the selected reducts. The results are average values of tenfold cross validation of C4.5 and Naive Bayes. 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. ρ-Value indicates the probability associated with the t-test. The smaller the value, the more significant the difference between the two average values is. What is 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 IV shows the classify performance by C4.5. Table IV 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 that obtain average classification accuracies over 80% except for the full attribute set. Table V shows the classification performance by the Naive Bayes classifier. From the table, we can get similar conclusion to Table IV. RMDPS and WRMDPS are still the methods that get the best average performance on all the datasets. On the whole, one may conclude that the presented methods outperform the compared algorithms and the full attribute set. In the last rows of Tables IV and V, 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, the 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 VI. 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,

TABLE III
COMPARISON OF THE WHOLE RUNNING TIME (SECONDS)

Algorithms	alma	anneal	clean	colon	credit	ecoli	glass	hepat	horse	ionos	lymph	segme	AverageTime
RMDPS	0.964	1.437	1.334	0.136	0.832	0.018	0.011	0.134	0.067	0.109	0.005	6.944	0.999
WRMDPS	0.884	1.467	1.365	0.156	0.808	0.018	0.008	0.143	0.070	0.109	0.005	7.035	1.006
L-FRFS	235.571	13.737	44.677	16.479	5.202	0.051	0.038	41.546	0.741	0.231	0.055	8.788	30.592
FDM	1.428	4.119	4.408	0.22	1.923	0.022	0.012	0.171	0.222	0.135	0.011	8.371	1.754
SPS	0.77	1.752	1.394	0.142	1.035	0.042	0.018	0.132	0.093	0.13	0.009	9.296	1.234
NFRS	69.671	6.975	206.063	49.36	4.18	0.062	0.037	2.907	0.465	2.083	0.018	37.515	31.611

TABLE IV
ACCURACY RESULTS OF ALGORITHMS WITH C4.5 CLASSIFIERS

Datasets	RMDPS	WRMDPS		$L-$FRFS		FDM		SPS		NFRS		fullset	
	Acc $\pm$ Std	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value
Amla	94.11 ± 0.21	94.11 ± 0.21	1.00○	94.55 ± 0.48	0.01+	93.06 ± 0.0	0.00−	66.91 ± 0.83	0.00−	92.52 ± 0.51	0.00−	81.33 ± 0.7	0.00−
Anneal	98.53 ± 0.06	98.53 ± 0.06	1.00○	98.18 ± 0.1	0.00−	98.51 ± 0.06	0.10○	98.53 ± 0.06	1.00○	88.41 ± 0.15	0.00−	98.52 ± 0.04	0.96○
Clean	77.09 ± 1.26	80.11 ± 1.5	0.01+	76.97 ± 0.92	0.66○	74.8 ± 0.24	0.00−	78.8 ± 0.74	0.02+	79.83 ± 0.36	0.00+	83.0 ± 0.49	0.00+
Colon	80.92 ± 0.6	80.52 ± 0.57	0.17○	87.01 ± 0.83	0.00+	79.91 ± 0.49	0.00−	71.49 ± 0.78	0.00−	79.53 ± 1.27	0.00−	82.73 ± 1.19	0.00+
Credit	72.77 ± 0.34	72.66 ± 0.36	0.02−	72.13 ± 0.28	0.00−	72.13 ± 0.28	0.00−	72.98 ± 0.21	0.01+	68.61 ± 0.24	0.00−	71.15 ± 0.19	0.00−
Ecoli	82.84 ± 0.32	82.84 ± 0.32	1.00○	82.84 ± 0.32	1.00○	82.84 ± 0.32	1.00○	82.84 ± 0.32	1.00○	82.84 ± 0.32	1.00○	82.84 ± 0.32	1.00○
Glass	68.24 ± 0.63	68.24 ± 0.63	1.00○	68.24 ± 0.63	1.00○	67.17 ± 1.12	0.01−	68.24 ± 0.63	1.00○	67.03 ± 0.77	0.00−	68.24 ± 0.63	1.00○
Hepat	90.21 ± 0.7	73.42 ± 2.11	0.00−	86.24 ± 0.91	0.00−	86.24 ± 0.91	0.00−	66.05 ± 1.29	0.00−	71.82 ± 3.45	0.00−	53.79 ± 1.59	0.00−
Horse	84.65 ± 0.08	84.65 ± 0.08	1.00○	84.81 ± 0.15	0.00+	84.95 ± 0.41	0.03+	85.08 ± 0.61	0.07○	68.65 ± 0.19	0.00−	83.47 ± 0.29	0.00−
Ionos	90.37 ± 0.6	90.37 ± 0.6	1.00○	64.1 ± 0.0	0.00−	80.91 ± 0.01	0.00−	91.35 ± 0.36	0.00+	89.8 ± 0.21	0.00−	89.68 ± 0.33	0.01−
Lymph	74.3 ± 0.35	74.3 ± 0.35	1.00○	74.3 ± 0.35	1.00○	74.3 ± 0.35	1.00○	73.12 ± 0.53	0.00−	67.72 ± 0.33	0.00−	76.47 ± 0.97	0.00+
Segme	96.83 ± 0.06	96.83 ± 0.06	1.00○	14.29 ± 0.0	0.00−	52.64 ± 0.08	0.00−	96.82 ± 0.07	0.35○	96.21 ± 0.1	0.00−	96.83 ± 0.05	0.44○
Average(Acc%)	84.23	83.03		75.3		78.94		79.35		79.39		80.66	
Lose/Win/Tie		2/1/9/		5/3/4/		8/1/3/		4/3/5/		10/1/1/		5/3/4/	

TABLE V
ACCURACY RESULTS OF ALGORITHMS WITH NAIVE BAYES CLASSIFIERS

Datasets	RMDPS	WRMDPS		$L-$FRFS		FDM		SPS		NFRS		fullset	
	Acc $\pm$ Std	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value	Acc $\pm$ Std	ρ-Value
Amla	94.65 ± 0.71	94.65 ± 0.71	1.00○	96.24 ± 0.39	0.00+	96.39 ± 0.24	0.00+	70.69 ± 0.26	0.00−	99.85 ± 0.12	0.00+	99.32 ± 0.28	0.00+
Anneal	87.71 ± 0.04	87.71 ± 0.04	1.00○	84.16 ± 0.06	0.00−	87.72 ± 0.05	0.50○	87.71 ± 0.04	1.00○	39.54 ± 0.07	0.00−	86.52 ± 0.07	0.00−
Clean	67.75 ± 0.2	73.81 ± 0.28	0.00+	69.57 ± 0.18	0.00+	70.17 ± 0.3	0.00+	69.63 ± 0.22	0.00+	76.15 ± 0.13	0.00+	74.1 ± 0.19	0.00+
Colon	85.48 ± 0.0	83.6 ± 0.31	0.00−	77.42 ± 0.31	0.00−	79.87 ± 0.89	0.00−	65.29 ± 0.45	0.00−	84.35 ± 0.41	0.00−	55.96 ± 0.49	0.00−
Credit	74.76 ± 0.27	74.66 ± 0.2	0.00−	74.0 ± 0.12	0.00−	74.0 ± 0.12	0.00−	74.69 ± 0.19	0.08○	70.92 ± 0.06	0.00−	75.29 ± 0.11	0.00+
Ecoli	85.52 ± 0.1	85.52 ± 0.1	1.00○	85.47 ± 0.11	0.00−	85.47 ± 0.11	0.00−	85.52 ± 0.1	1.00○	85.52 ± 0.1	1.00○	85.52 ± 0.1	1.00○
Glass	48.87 ± 0.69	48.87 ± 0.69	1.00○	48.87 ± 0.69	1.00○	43.2 ± 0.54	0.00−	48.87 ± 0.69	1.00○	48.47 ± 0.54	0.00−	48.87 ± 0.69	1.00○
Hepat	82.48 ± 1.68	87.54 ± 1.02	0.00+	84.68 ± 2.37	0.00+	84.68 ± 2.37	0.00+	59.06 ± 1.52	0.00−	93.61 ± 0.58	0.00+	74.87 ± 0.48	0.00−
Horse	81.02 ± 0.1	81.02 ± 0.1	1.00○	81.38 ± 0.3	0.00+	79.36 ± 0.21	0.00−	78.85 ± 0.17	0.00−	59.55 ± 0.24	0.00−	77.17 ± 0.11	0.00−
Ionos	82.16 ± 0.08	82.16 ± 0.08	1.00○	64.1 ± 0.0	0.00−	76.37 ± 0.04	0.00−	85.68 ± 0.26	0.00+	86.97 ± 0.12	0.00+	82.48 ± 0.12	0.00+
Lymph	80.01 ± 0.18	80.01 ± 0.18	1.00○	80.01 ± 0.18	1.00○	80.01 ± 0.18	1.00○	79.38 ± 0.45	0.01−	72.18 ± 0.32	0.00−	82.77 ± 0.2	0.00+
Segme	80.48 ± 0.03	80.48 ± 0.03	1.00○	14.29 ± 0.0	0.00−	47.64 ± 0.05	0.00−	80.65 ± 0.07	0.00+	86.0 ± 0.06	0.00+	79.88 ± 0.03	0.00−
Average(Acc%)	79.24	80.00		71.69		75.42		73.84		75.26		76.90	
Lose/Win/Tie		2/2/8/		6/4/2/		7/3/2/		5/3/4/		6/5/1/		5/5/2/	

TABLE VI
RUNNING TIME RESULTS OF THE FIRST STEP (SECONDS)

Algorithms	alma	anneal	clean	colon	credit	ecoli	glass	hepat	horse	ionos	lymph	segme	AverageTime
RMDPS	0.823	1.365	1.186	0.108	0.75	0.013	0.008	0.105	0.055	0.096	0.003	6.146	0.888
WRMDPS	0.699	1.384	1.156	0.12	0.706	0.013	0.005	0.103	0.054	0.091	0.003	6.038	0.864
SPS	0.616	1.415	1.158	0.115	0.729	0.013	0.005	0.103	0.055	0.096	0.003	6.416	0.894

TABLE VII
RUNNING TIME RESULTS OF THE SECOND STEP (SECONDS)

Algorithms	alma	anneal	clean	colon	credit	ecoli	glass	hepat	horse	ionos	lymph	segme	AverageTime
RMDPS	0.087	0.05	0.08	0.017	0.051	0.003	0.002	0.017	0.007	0.007	0.001	0.432	0.063
WRMDPS	0.078	0.049	0.08	0.014	0.05	0.003	0.002	0.016	0.007	0.007	0.001	0.536	0.07
SPS	0.124	0.144	0.139	0.022	0.178	0.01	0.005	0.025	0.022	0.017	0.004	0.924	0.1345

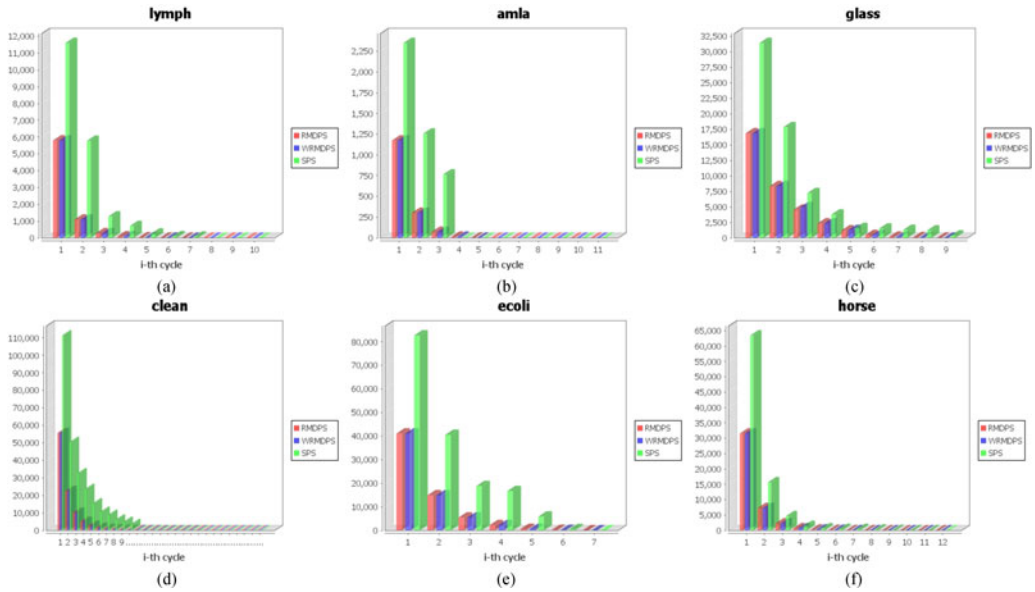

Fig. 3. Number of remaining object pairs in the ith cycle.

TABLE VIII
RUNNING TIME RESULTS OF THE THIRD STEP (SECONDS)

Algorithms	alma	anneal	clean	colon	credit	ecoli	glass	hepat	horse	ionos	lymph	segme	AverageTime
RMDPS	0.054	0.022	0.068	0.011	0.03	0.002	0.002	0.024	0.009	0.011	0.001	0.367	0.048
WRMDPS	0.106	0.035	0.13	0.022	0.052	0.002	0.002	0.024	0.009	0.011	0.001	0.461	0.071
SPS	0.029	0.193	0.097	0.005	0.129	0.02	0.008	0.004	0.016	0.017	0.002	1.956	0.206

from Table VII, 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 VIII, 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 uncovered 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 uncovered 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 uncovered pairs are large at each iteration.

VI. 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 the 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 the Davis–Logemann–Loveland-based strategy [44], the 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 the fuzzy rough framework from the viewpoint of the object pair. We think the presented study can supply an optional angle to consider the attribute reduction issue, since most of the existing attribute selection algorithms mainly take the angle of the attribute set. However, one should notice that there are also improved and optimized approaches for attribute reduction from the viewpoint of the attribute set, such as efficient positive-region-based approaches [47]–[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 the 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 the object pair.

APPENDIX

As shown in Algorithms 1 and 2, the While circulation should run exactly $|\text{Red}|+1$ cycles. Suppose maxNum_i is the value of maxNum in the ith cycle, selAtt_i represents the selected attribute in the ith selection, and MDP'_i represents the number of remaining $\text{MDP}'_C D$. Note that $\text{maxNum}_0 = \text{maxNum}_{|\text{Red}|+1} = 0$, $\text{MDP}'_{|\text{Red}|+1} = 0$. Then, we have

$$\begin{aligned}\text{maxNum}_1 + \text{maxNum}_2 + \cdots + \text{maxNum}_{|\text{Red}|} &= |\text{MDP}'_C D(U)| \quad (56)\\ \text{MDP}'_i = |\text{MDP}'_C D(U)| - (\text{maxNum}_1 + \cdots &+ \text{maxNum}_{i-1}), i \geq 1. \quad (57)\end{aligned}$$

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

$$\text{MDP}'_i = \text{MDP}'_1 * \partial^{i-1}. \quad (58)$$

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

$$\begin{aligned}\text{MDP}'_{|\text{Red}|+1} = 0 &\Leftrightarrow 0 \leq \text{MDP}'_{|\text{Red}|+1} < 1\\ &\Leftrightarrow 0 \leq \text{MDP}'_1 * \partial^{|\text{Red}|} < 1 \quad (59)\end{aligned}$$

which implies

$$\partial < \sqrt[|\text{Red}|]{\frac{1}{\text{MDP}'_1}} = \sqrt[|\text{Red}|]{\frac{2}{|U|(|U|-1)}}. \quad (60)$$

Here, we choose $\partial = \sqrt[|\text{Red}|]{\frac{2}{|U|(|U|-1)}}$ in a most conservative approach. Obviously, its range is $(0, 1)$. So

$$\begin{aligned}&\text{MDP}'_1 + \text{MDP}'_2 + \cdots + \text{MDP}'_{|Red|+1}\\ &= \text{MDP}'_1 * \partial^0 + \text{MDP}'_1 * \partial^1 + \cdots + \text{MDP}'_1 * \partial^{|Red|}\\ &= \text{MDP}'_1 * \frac{1-\partial^{|\text{Red}|+1}}{1-\partial}\\ &< \text{MDP}'_1 * \frac{1}{1-\partial}\\ &= \frac{|U|(|U|-1)}{2} * \frac{1}{1-\partial}\\ &< \frac{|U|^2}{2} * \frac{1}{1-\partial} \quad (61)\end{aligned}$$

which means the time complexity for steps 3–18 is $O(\frac{|U|^2|C|}{2} * \frac{1}{1-\partial})$, in which

$$\partial = \sqrt[|\text{Red}|]{\frac{2}{|U|(|U|-1)}} \quad (62)$$

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 $|\text{Red}|$ represents the number of attributes contained in the reduct with respect to the given decision table. Then, $\frac{|U|^2|C|}{2} * \frac{1}{1-\partial} \leq |U|^2|C| \Leftrightarrow |U|(|U|-1) \geq 2^{|\text{Red}|+1}$.

Proof: It can be easily proved using (62). ■

REFERENCES

[1] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982.

[2] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Norwell, MA, USA: Kluwer, 1991.

[3] J. Dai and Q. Xu, "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 211–221, 2013.

[4] Q. Hu, W. Pedrycz, D. Yu, and J. Lang, "Selecting discrete and continuous features based on neighborhood decision error minimization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 1, pp. 137–150, Feb. 2010.

[5] P. Maji, "A rough hypercuboid approach for feature selection in approximation spaces," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 16–29, Jan. 2014.

[6] H. Chen, T. Li, C. Luo, S.-J. Horng, and G. Wang, "A decision-theoretic rough set approach for dynamic data mining," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 1958–1970, Dec. 2015.

[7] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web categorization," *Fuzzy Sets Syst.*, vol. 141, no. 3, pp. 469–485, 2004.

[8] Q. Hu, S. An, and D. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection," *Inf. Sci.*, vol. 180, no. 22, pp. 4384–4400, 2010.

[9] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, nos. 2/3, pp. 191–209, 1990.

[10] W. Z. Wu, J. S. Mi, and W. X. Zhang, "Generalized fuzzy rough sets," *Inf. Sci.*, vol. 151, pp. 263–282, 2003.

[11] J. Liang, F. Wang, C. Dang, and Y. Qian, "A group incremental approach to feature selection applying rough set technique," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 294–308, Feb. 2014.

[12] Y. Qian, J. Liang, D. Li, H. Zhang, and C. Dang, "Measures for evaluating the decision performance of a decision table in rough set theory," *Inf. Sci.*, vol. 178, no. 1, pp. 181–202, 2008.

[13] Y. Qian, J. Liang, Y. Yao, and C. Dang, "MGRS: A multi-granulation rough set," *Inf. Sci.*, vol. 180, no. 6, pp. 949–970, 2010.

[14] J. Dai, W. Wang, and Q. Xu, "An uncertainty measure for incomplete decision tables and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1277–1289, Aug. 2013.
[15] J. Dai, Q. Hu, J. Zhang, H. Hu, and N. Zheng, "Attribute selection for partially labeled categorical data by rough set approach," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2460–2471, Sep. 2017.
[16] C. Z. Wang *et al.*, "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.
[17] M. De Cock, C. Cornelis, and E. E. Kerre, "Fuzzy rough sets: The forgotten step," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 121–130, Feb. 2007.
[18] D. S. Yeung, D. Chen, E. C. Tsang, J. W. Lee, and X. Wang, "On the generalization of fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343–361, Jun. 2005.
[19] E. C. Tsang, C. Wang, D. Chen, C. Wu, and Q. Hu, "Communication between information systems using fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 3, pp. 527–540, Jun. 2013.
[20] R. Jensen and C. Cornelis, "Fuzzy-rough nearest neighbour classification," in *Transactions on Rough Sets XIII*. New York, NY, USA: Springer, 2011, pp. 56–72.
[21] P. Maji, "Fuzzy-rough supervised attribute clustering algorithm and classification of microarray data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 222–233, Feb. 2011.
[22] B. Huang, D. Wei, H. Li, and Y. Zhuang, "Using a rough set model to extract rules in dominance-based interval-valued intuitionistic fuzzy information systems," *Inf. Sci.*, vol. 221, pp. 215–229, 2013.
[23] Q. He, C. Wu, D. Chen, and S. Zhao, "Fuzzy rough set based attribute reduction for information systems with fuzzy decisions," *Knowl.-Based Syst.*, vol. 24, no. 5, pp. 689–696, 2011.
[24] Q. Hu, L. Zhang, S. An, D. Zhang, and D. Yu, "On robust fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 636–651, Aug. 2012.
[25] D. Chen and Y. Yang, "Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1325–1334, Oct. 2014.
[26] S. Zhao, H. Chen, C. Li, X. Du, and H. Sun, "A novel approach to building a robust fuzzy rough classifier," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 769–786, Aug. 2015.
[27] J. Dai and H. Tian, "Fuzzy rough set model for set-valued data," *Fuzzy Sets Syst.*, vol. 229, pp. 54–68, 2013.
[28] Q. Shen and R. Jensen, "Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring," *Pattern Recognit.*, vol. 37, no. 7, pp. 1351–1363, 2004.
[29] R. B. Bhatt and M. Gopal, "On fuzzy-rough sets approach to feature selection," *Pattern Recognit. Lett.*, vol. 26, no. 7, pp. 965–975, 2005.
[30] N. M. Parthalain, R. Jensen, and Q. Shen, "Fuzzy entropy-assisted fuzzy-rough feature selection," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2006, pp. 423–430.
[31] Q. Hu, D. Yu, and Z. Xie, "Information-preserving hybrid data reduction based on fuzzy-rough techniques," *Pattern Recognit. Lett.*, vol. 27, no. 5, pp. 414–423, 2006.
[32] R. Jensen and Q. Shen, "Fuzzy-rough sets assisted attribute selection," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 73–89, Feb. 2007.
[33] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Aug. 2009.
[34] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 385–389, Apr. 2012.
[35] C. Liu, Q. Cao, and G. Zheng, "New definition of dependency in fuzzy rough feature selection," in *Proc. Int. Conf. Syst. Sci., Eng. Des. Manuf. Informatization*, 2011, vol. 2, pp. 271–274.
[36] Y. Qian, Q. Wang, H. Cheng, J. Liang, and C. Dang, "Fuzzy-rough feature selection accelerator," *Fuzzy Sets Syst.*, vol. 258, pp. 61–78, 2015.
[37] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," *Intell. Decision Support Theory Decision Library*, vol. 11, pp. 331–362, 1992.
[38] D. Dubois and H. Prade, "Putting rough sets and fuzzy sets together," in *Intelligent Decision Support*. New York, NY, USA: Springer, 1992, pp. 203–232.
[39] J. Fernandez Salido and S. Murakami, "Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations," *Fuzzy Sets Syst.*, vol. 139, no. 3, pp. 635–660, 2003.
[40] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 126, no. 2, pp. 137–155, 2002.
[41] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.
[42] J. Dai and Y. Li, "A attribute reduction algorithm based on the decision system of rough set," *Mini-Micro Comput. Syst.*, vol. 24, no. 3, pp. 523–526, 2003.
[43] N. Iizuka *et al.*, "Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection," *Lancet*, vol. 361, no. 9361, pp. 923–929, 2003.
[44] R. Jensen, A. Tuson, and Q. Shen, "Finding rough and fuzzy-rough set reducts with SAT," *Inf. Sci.*, vol. 255, pp. 100–120, 2014.
[45] J. Dai, W. Wang, H. Tian, and L. Liu, "Attribute selection based on a new conditional entropy for incomplete decision systems," *Knowl.-Based Syst.*, vol. 39, pp. 207–213, 2013.
[46] J. Dai, H. Han, Q. Hu, and M. Liu, "Discrete particle swarm optimization approach for cost sensitive attribute reduction," *Knowl.-Based Syst.*, vol. 102, pp. 116–126, 2016.
[47] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: An accelerator for attribute reduction in rough set theory," *Artif. Intell.*, vol. 174, pp. 597–618, 2010.
[48] J. Dai, Q. Hu, H. Hu, and D. Huang, "Neighbor inconsistent pair selection for attribute reduction by rough set approach," *IEEE Trans. Fuzzy Syst.*, 2017, doi: 10.1109/TFUZZ.2017.2698420, to be published.
[49] J. Fan, Y. Jiang, and Y. Liu, "Quick attribute reduction with generalized indiscernibility models," *Inf. Sci.*, vols. 397/398, pp. 15–36, 2017.

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, China. He has authored or coauthored more than 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 Hefei University of Technology, Hefei, China, in 2015. He is currently working toward 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.

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. and Ph.D. degrees in computers with applications from Shanxi University, Taiyuan, China, in 2005 and 2011, respectively.

He is currently a Professor with the School of Computer and Information Technology, Shanxi University. His current research interests include pattern recognition, feature selection, rough set theory, granular computing, and artificial intelligence.

Debiao Huang received the B.Sc. degree in computer science and technology from Northeast University, Shenyang, China, in 2013. He is currently working toward the Graduate degree with the College of Computer Science, Zhejiang University, Hangzhou, China.

His research interests include machine learning, data mining, and rough sets.