Incremental Perspective for Feature Selection Based on Fuzzy Rough Sets

Yanyan Yang, Degang Chen, Hui Wang, and Xizhao Wang, *Fellow, IEEE*

***Abstract*—Feature selection based on fuzzy rough sets is an effective approach to select a compact feature subset that optimally predicts a given decision label. Despite being studied extensively, most existing methods of fuzzy rough set based feature selection are restricted to computing the whole dataset in batch, which is often costly or even intractable for large datasets. To improve the time efficiency, we investigate the incremental perspective for fuzzy rough set based feature selection assuming data can be presented in sample subsets one after another. The key challenge for the incremental perspective is how to add and delete features with the subsequent arrival of sample subsets. We tackle this challenge with strategies of adding and deleting features based on the relative discernibility relations that are updated as subsets arrive sequentially. Two incremental algorithms for fuzzy rough set based feature selection are designed based on the strategies. One updates the selected features as each sample subset arrives, and outputs the final feature subset where no sample subset is left. The other updates the relative discernibility relations but only performs feature selection where there is no further subset arriving. Experimental comparisons suggest our incremental algorithms expedite fuzzy rough set based feature selection without compromising performance.**

***Index Terms*—Attribute reduction, feature selection, fuzzy rough sets, incremental learning, relative discernibility relation.**

I. Introduction

FEATURE selection is a commonly used preprocessing step in machine learning, data mining, and pattern recognition. It is a process of selecting a compact and informative subset from the original features that can be used for building a satisfactory predictive model [1], [2]. In recent years, data have become increasingly larger and larger in both size and dimensionality, which poses serious challenges for most of the existing feature selection algorithms. Therefore, efficient feature selection from large datasets is one such challenge.

The fuzzy rough set theory [3], [4] provides the basis for an effective approach to feature selection through the use of a fuzzy similarity relation, which describes the similarity between pairs of data samples (i.e., instances). Fuzzy rough set is mainly used in classification to address the inconsistency between features and decision labels, i.e., some samples have the similar feature values but different decision labels. The inconsistency can be measured by using the lower approximation in fuzzy rough sets to assign a membership to every sample with respect to decision labels. By keeping the membership of every sample unchanged, fuzzy rough set based feature selection (FS-FRS), usually called attribute reduction, can remove redundant or irrelevant features to find an informative feature subset [6].

Despite the extensive investigation in the literature, most existing methods of FS-FRS are restricted to the batch processing, which handles all samples of a dataset in batch mode all at once. Quite often, this is uneconomic, and even impractical for large datasets that easily exceed the memory capacity. This reveals one weakness of those batch algorithms in terms of the runtime. New feature selection algorithms are thus needed that scale well with the increase of data size [7], [8]. Incremental feature selection has been explored recently to deal with the case in which data arrive sequentially (that is dynamic) or a large dataset (due to its big size) has to be cut into small subsets that are then presented sequentially. There are some state-of-the-art methods for incremental feature selection based on rough sets. Although these incremental methods are more efficient than batch feature selection methods based on rough sets, they do not provide an essential insight into the incremental mechanism of FS-FRS from the viewpoint of the successive arrival of sample subsets.

Motivated by the above observations, we take an incremental approach where a real-valued dataset is divided into a sequence of sample subsets that are added in succession, and each sample subset is sequentially processed upon its arrival. In the incremental approach, we present the strategies of adding and deleting features based on the relative discernibility relations that are updated as sample subsets arrive continuously. Based on the strategies, we design two incremental algorithms for FS-FRS: 1) updating the selected feature subset where each sample subset arrives, and 2) only performing feature selection where no sample subset is left. In the first version, upon the arrival of each subset, the relative discernibility relations of each feature and the feature set are incrementally computed to update the selected feature subset. When there is no further

Manuscript received October 13, 2016; revised March 3, 2017; accepted June 3, 2017. Date of publication June 27, 2017; date of current version May 31, 2018. This work was supported in part by the NSFC under Grant 71471060, Grant 61170040, Grant 71371063, in part by the JCYJ under Grant 20150324140036825, in part by the Ulster University's Research Challenge Fund under Grant 70595Q, and in part by Fundamental Research Funds for the Central Universities (2018ZD06). *(Corresponding author: Yanyan Yang.)*

Y. Yang is with the School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China (e-mail: hrsinbba@163.com).

D. Chen is with the Department of Mathematics and Physics, North China Electric Power University, Beijing 102206, China (e-mail: wscz973@163.com).

H. Wang is with the School of Computing and Mathematics, University of Ulster, Coleraine BT52 1SA, U.K. (e-mail: h.wang@ulster.ac.uk).

X. Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: xizhaowang@ieee.org).

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.2017.2718492

sample subset arriving, the feature subset will become the final one for the whole dataset. In the second version, we update the relative discernibility relations of each feature and the feature set every time a sample subset arrives. When no further subset arrives, the obtained relative discernibility relations are used to search for the final feature subset. Theoretically, it is clear that the second algorithm is more efficient than the first one as it does not update the feature subset every time a sample subset arrives. This fact will be validated in our experimental results.

The unique contributions that distinguish this paper from the existing batch feature selection methods and incremental feature selection methods in the framework of rough sets are threefold.

1) Our work integrates the incremental fashion into FS-FRS from a dataset.
2) The strategies of adding and deleting features are developed based on the updated relative discernibility relations of each feature and the feature set.
3) Two efficient incremental algorithms for FS-FRS are presented to start with an empty set to compute a reduct from the dataset, with extensive experimental comparisons.

The remainder of the paper is organized as follows. Related works are discussed in Section II. In Section III, some preliminaries on fuzzy rough sets are presented. In Section IV, the relative discernibility relation based algorithm for feature selection is presented based on fuzzy rough sets. The incremental algorithms of FS-FRS are presented in Section V. In Section VI, experimental results demonstrate the time efficiency of our proposed incremental algorithms without sacrificing the quality of selected features. This paper is concluded with a summary in Section VII.

II. Related Works

Our work to be presented in this paper is closely related to the following studies: The batch methods of FS-FRS, rough set based incremental rules, rough set based incremental approximations, and rough set based incremental feature selection. In the following sections, we review some important related works in terms of the above four aspects.

A. Batch Methods of FS-FRS

Many efforts have been made to investigate FS-FRS. As a pioneering work on FS-FRS, Jesen and Shen [9] proposed a dependence function based heuristic algorithm to find a reduct. Other research works on FS-FRS mostly improve the method in [9]. For example, Bhatt *et al.* [10] defined a compact domain to improve the time efficiency of the algorithm in [9]. Hu *et al.* [11] proposed an information entropy based algorithm for feature selection with fuzzy rough sets. In [12], a fuzzy extension to crisp discernibility matrices was proposed to search for a feature subset. It has been noted in [13] that these heuristic algorithms cannot find a proper reduct but an overreduct or underreduct due to their stop criteria. To find proper reducts, Tsang *et al.* [5] introduced the discernibility matrix based approach to fuzzy rough sets, which requires heavy computational load since it computes and searches for every element in the discernibility matrix. To improve the time efficiency in [5], Chen *et al.* [13] developed the sample pair selection method to search for all minimal elements in the discernibility matrix, and only employed all minimal elements to find reducts of a fuzzy decision table. Wang *et al.* [14] proposed a fitting model for FS-FRS to better reflect the classification ability of a selected feature subset.

Despite the extensive investigations in the literature, the works discussed above are restricted to the batch computing, which handles all samples of a dataset all at once. Such batch methods are often costly, and even intractable for large datasets that easily exceed the memory capacity. This inevitably poses great challenges to traditional batch methods of FS-FRS. As an efficient technique, the incremental approach has been introduced into rough sets to update knowledge such as decision rules, approximations, feature selection from dynamic datasets in which samples, features or feature values vary with time.

B. Rough Set Based Incremental Decision Rules

There are several works for updating decision rules from dynamic datasets. Fan *et al.* [15] proposed an incremental method for updating decision rules by analyzing different cases of strength index change. Shan *et al.* [16] updated decision rules by updating the decision matrices and decision functions at the arrival of a sample. A rule induction method was proposed in [17] based on fuzzy rough sets when samples vary. Zheng *et al.* [18] investigated a tree-based method of updating decision rules with the arrival of samples. Blaszczynski and Slowinski [19] proposed an algorithm for updating decision rules in terms of adding samples by extending the Apriori algorithm to the variable consistency dominance based rough set. Tripathy *et al.* [20] proposed an enhanced rule induction algorithm (ELEM)-based algorithm to update decision rules with the arrival of a sample. Tsumoto [21] proposed an algorithm for updating decision rules based on the accuracy and coverage with the arrival of a sample. When coarsening and refining attribute values, Chen *et al.* [22] investigated the dynamic maintenance of decision rules.

C. Rough Set Based Incremental Approximations

Many researchers have focused on incrementally updating approximations based on rough sets with the variations of the feature set, the feature value, and the sample set, respectively.

With the variation of the feature set, Chan [23] presented an incremental method for updating approximations. Li *et al.* [24] proposed methods to update approximations of the characteristic relation based rough sets by the variations of the upper and lower boundary regions. Cheng [25] proposed two incremental methods for fast computing the rough fuzzy approximation based on the boundary set and the cut sets of a fuzzy set, respectively. Zhang *et al.* [26] proposed an incremental approach for updating approximations of set-valued information systems based on the probabilistic neighborhood rough set model. Li *et al.* [27] proposed an incremental algorithm for updating approximations based on P-generalized decision of a sample. Luo *et al.* [28] focused on dynamically maintaining approximations in set-valued ordered decision systems. Yang *et al.* [29] updated multigranulation rough approximations with the increasing granular structures. Liu *et al.* [30] investigated the incremental method of updating approximations in probabilistic rough sets.

Chen *et al.* [31] proposed an incremental algorithm for approximations of a concept when coarsening or refining feature values. In incomplete ordered decision systems, Chen *et al.* [32] presented a method to dynamically maintain approximations of upward and downward unions when feature values change. Luo *et al.* [33] proposed two incremental algorithms for computing rough approximations with respect to the addition and removal of criteria values.

With varying the sample set, Chen *et al.* [34] designed incremental methods for updating approximations based on variable precision rough set model. Zhang *et al.* [35] proposed a method of updating approximations based on neighborhood rough sets. Li *et al.* [36] proposed a dynamic maintenance approach for updating approximations in the ordered information system. Luo *et al.* [37] presented two incremental algorithms for computing approximations in disjunctive/conjunctive set-valued information systems. Zhang *et al.* [38] developed a novel matrix-based algorithm for fast updating approximations in dynamic composite information systems. Zeng *et al.* [39] proposed an incremental approach for updating approximations of Gaussian kernelized fuzzy rough sets with the variation of the sample set. Luo *et al.* [40] proposed incremental algorithms for updating approximations in decision theoretic rough sets with respect to the addition and deletion of samples.

D. Rough Set Based Incremental Feature Selection

Many attempts have been made to investigate rough set based incremental feature selection with the variations of the feature set, the feature values, and the sample set, respectively.

When adding features, Wang *et al.* [41] developed a dimension (i.e., feature) incremental strategy for feature subset based on the updating mechanisms of three measures of information entropy. With adding and deleting features, Shu *et al.* [42] proposed two algorithms for updating feature selection by the incremental computation of the positive region in incomplete decision systems. As the feature set varies dynamically, an incremental feature selection algorithm was proposed in [43] by updating knowledge granulation in the set-valued information system. When adding and deleting features, Zeng *et al.* [44] employed the dependence function to analyze the incremental mechanisms for feature selection based on Gaussian kernelized fuzzy rough sets in hybrid information systems.

With dynamically varying feature values, Wang *et al.* [45] developed an incremental algorithm for feature selection based on the incremental computation of three representative information entropies. For single sample and multiple samples with varying feature values, Shu *et al.* [46] developed two incremental feature selection algorithms based on the incremental computation of the positive region.

With the arrival of a sample, Liu [47] proposed an incremental feature selection algorithm to find the minimal reduct from an information system without decision labels. When adding a new sample into a decision table, Hu *et al.* [48] proposed an incremental feature selection based on the positive region. Afterward, Hu *et al.* [49] proposed an incremental algorithm for finding all reducts based on the modified discernibility matrix. At the arrival of a new sample, Yang [50] proposed an incremental feature selection algorithm by updating the discernibility matrix. At the arrival of a sample, Chen *et al.* [51] proposed an incremental algorithm for feature selection based on variable precision rough sets by the strategies of adding and deleting features. As a sample with real-valued features arrives, Li *et al.* [60] investigated the incremental mechanisms for feature selection based on neighborhood rough sets. When a group of samples arrives, Liang *et al.* [52] developed a group incremental algorithm for feature selection based on the investigation of incremental mechanisms for three measures of information entropy including Shannon's entropy, complementary entropy, and combination entropy. To address the time/space complexity issue of the current incremental feature selection algorithms, Yang *et al.* [53] presented a novel incremental algorithm for rough set based feature selection by integrating an active sample selection process that discards useless incoming samples and selects useful incoming samples into the feature selection process that determines how to add and delete features in the current selected feature subset.

From what have been discussed above, FS-FRS has not yet been studied incrementally from the viewpoint of the sample subset sequence, where a dataset is divided into a sequence of sample subsets to be added in succession. This motivates our study in this paper.

III. Preliminaries

This section reviews fuzzy logical operators, fuzzy approximation operators and FS-FRS.

A. Fuzzy Logical Operators

In this section, we present and exemplify five fuzzy logical operators [3], [54], namely t- norm, t- conorm, negator, dual, residual implication and its dual operation.

t- norm is a function $T : [0,1] \times [0,1] \rightarrow [0,1]$ satisfying the following:

1) commutativity: $T(x,y) = T(y,x)$;
2) associativity: $T(T(x,y),z) = T(x,T(y,z))$;
3) monotonicity: $x \leq \alpha, y \leq \beta, T(x,y) \leq T(\alpha,\beta)$;
4) boundary condition: $T(x,1) = T(1,x) = x$.

The most popular continuous t- norms include the standard min operator $T_M(x,y) = \min\{x,y\}$, the algebraic product $T_P(x,y) = x \cdot y$, and the Lukasiewicz t-norm $T_L(x,y) = \max\{0, x+y-1\}$.

t-conorm is an increasing, commutative, and associative function $S : [0,1] \times [0,1] \rightarrow [0,1]$ satisfying the boundary condition $\forall x \in [0,1],\ S(x,0) = x$. The well-known continuous t-conorms include the standard max operator $S_M(x,y) = \max\{x,y\}$, the probabilistic sum $S_P(x,y) = x + y - x \cdot y$, and the bounded sum $S_L(x,y) = \min\{1, x+y\}$.

A negator N is a decreasing function $N : [0,1] \rightarrow [0,1]$ satisfying $N(0) = 1$ and $N(1) = 0$. $N_S(x) = 1 - x$ is called the standard negator. A negator N is called involutive if $N(N(x)) = x, \forall x \in [0,1]$. A t-norm T and a t-conorm S are called dual with respect to N iff $S(N(x), N(y)) = N(T(x, y))$ and $T(N(x), N(y)) = N(S(x, y))$.

Let $X : U \to [0, 1]$ be a fuzzy set and $F(U)$ be the fuzzy power set on U. For each $X \in F(U)$, the symbol $co_N X$ is denoted as the fuzzy complement of X determined by a negator N, i.e., for every $x \in U$, $(co_N X)(x) = N(N(x))$.

Given a lower semicontinuous triangular norm T, the T-residuated implication is a function $\vartheta : [0, 1] \times [0, 1] \to [0, 1]$ satisfying $\vartheta(x, y) = \sup\{z \in [0, 1] : T(x, z) \le y\}$ for $\forall x, y \in [0, 1]$. T-residuated implications include the Lukasiewicz implication ϑ_L based on $T_L : \vartheta_L(x, y) = \min\{1 - x + y, 1\}$.

Given an upper semicontinuous triangular conorm S, the dual of T-residuated implication with respect to N is a function $\sigma : [0, 1] \times [0, 1] \to [0, 1]$ that satisfies $\sigma(x, y) = \inf\{z \in [0, 1], S(x, z) \ge y\}$ for every $x, y \in [0, 1]$. ϑ and σ are dual in terms of N if $\forall x, y \in [0, 1]$, $\sigma(x, y) = N(\vartheta(N(x), N(y)))$ or $\vartheta(x, y) = N(\sigma(N(x), N(y)))$.

B. Fuzzy Rough Sets

Let U be a finite set of samples and R a binary relation on U, R is called a fuzzy T-similarity relation if for $\forall x, y, z \in U$, it satisfies reflexivity ($R(x, x) = 1$), symmetry ($R(x, y) = R(y, x)$), and T-transitive ($T(R(x, y), R(y, z)) \le R(x, z)$).

Fuzzy rough sets were introduced by Dubois and Prade [3], [4], and then studied in [58] and [59]. For $\forall X \in F(U)$, the fuzzy approximation operators can be summarized as follows:

$$\overline{R_T} X(x) = \sup_{u \in U} T(R(x, u), X(u));$$
$$\underline{R_S} X(x) = \inf_{u \in U} S(N(R(x, u)), X(u));$$
$$\overline{R_\sigma} X(x) = \sup_{u \in U} \sigma(N(R(x, u)), X(u));$$
$$\underline{R_\vartheta} X(x) = \inf_{u \in U} \vartheta(R(x, u), X(u)).$$

From the viewpoint of the granular computing, $(\overline{R_T}, \underline{R_\vartheta})$ and $(\overline{R_\sigma}, \underline{R_S})$ are two pairs of approximation operators, since they can be represented by their individual fuzzy granules: $\overline{R_T} x_\lambda$ and $\overline{R_\sigma} x_\lambda$. Based on the granular structure, Chen *et al.* [55] characterized FS-FRS. Due to space limitations, this paper only concerns about the pair $(\overline{R_T}, \underline{R_\vartheta})$, and it has been noted in [55] that the corresponding conclusions for the pair $(\overline{R_\sigma}, \underline{R_S})$ can be similarly obtained.

C. Fuzzy Rough Set Based Feature Selection

In this section, we mainly review FS-FRS, and the discernibility matrix based approach to obtain a selected feature subset [55].

In this paper, $(U, R \cup D)$ is used to represent a fuzzy decision system, where U is the universe of discourse, R is the set of real-valued features, and $D = \{d\}$ is the set of symbol decision feature. The subset $P \subseteq R$ can be represented by a fuzzy T-similarity relation $P(x, y)$ for $\forall x, y \in U$, and $P(x, y) = \min_{a \in P}(a(x, y))$. Where no confusion arises, we use the set R to represent its corresponding similarity relation in the rest of this paper.

Assume $U/D = \{[x]_D : x \in U\}$ is the decision partition of U, where $[x]_D = \{y \in U : d(x) = d(y)\}$ is called the decision class to which the sample x belongs. The membership function of the decision class $[x]_D$ is defined as $[x]_D(y) = \{1,\ y \in [x]_D; 0,\ y \notin [x]_D$. Thus, for $x \in U$, we have $\overline{R_T}[y]_D(x) = \sup_{u \in [y]_D} R(x, u)$ and $\underline{R_\vartheta}[y]_D(x) = \inf_{u \notin [y]_D} \vartheta(R(x, u), 0)$.

The positive region of R with respect to D is defined as $Pos_R(D)(x) = \bigcup_{X \in U/D} \underline{R_\vartheta} X(x)$ for every $x \in U$.

According to Yang *et al.* [56], $Pos_R(D)(x) = \underline{R_\vartheta}[x]_D(x)$ holds for $\forall x \in U$, which actually means that the membership degree of each sample belonging to the positive region gets its value at the lower approximation of its corresponding decision class. By preserving the positive region unchanged, FS-FRS is defined as follows.

Definition 1 ([56]): Let $(U, R \cup D)$ be a fuzzy decision table. A subset $P \subseteq R$ is called a reduct of R relative to D, if the following conditions are satisfied: 1) $Pos_P(D)(x) = Pos_R(D)(x)$ for $\forall x \in U$; 2) for $\forall a \in P$, there exists $y \in U$ satisfying $Pos_{P-\{a\}}(D)(y) < Pos_R(D)(y)$.

The first condition means that a reduct P can preserve the positive region. The second one implies that for $\forall a \in P$, $P - \{a\}$ cannot preserve the positive region. Hence, a reduct P is a minimal feature subset that keeps the positive region. As a commonly used procedure for reducts, the discernibility matrix based approach in [5] is briefly stated as follows.

Let $U = \{x_1, x_2, \ldots, x_n\}$, $M_D(U, R) = (c_{ij})_{n \times n}$ is called the discernibility matrix of $(U, R \cup D)$, if

$$c_{ij} = \begin{cases} \{a \in R : T(a(x_i, x_j), \lambda(x_i)) = 0\}, & d(x_i) \ne d(x_j); \\ \emptyset, & \text{otherwise} \end{cases}$$

is the discernibility feature set discerning (x_i, x_j). $\lambda(x_i) = Pos_R(D)(x_i) = \inf_{u \notin [x_i]_D} \vartheta(R(x_i, u), 0)$ is the membership degree of x_i belonging to the positive region.

$f_D(U, R) = \bigwedge_{\substack{c_{ij} \in M_D(U,R) \\ c_{ij} \ne \emptyset}} \{\bigvee c_{ij}\}$ is called the discernibility function of $(U, R \cup D)$. If $f_D(U, R) = \bigvee_{k=1}^{t} (\bigwedge A_k)$ is the minimal disjunctive normal form of $f_D(U, R)$, $Red_D(R) = \{A_1, \ldots, A_t\}$ is then the set of all the reducts. The intersection of all the reducts is denoted as $\mathrm{Core}_D(R) = \bigcap Red_D(R)$, which is called the relative core of $(U, R \cup D)$. However, it is sufficient to find one reduct in many real-world applications. The following theorems are the basis of finding a reduct.

Theorem 1 ([57]): $\mathrm{Core}_D(R) = \{a : c_{ij} = \{a\} \in M_D(U, R)\}$.

Theorem 2 ([57]): $P \subseteq R$ is a reduct of R iff the following conditions hold: 1) $P \bigcap c_{ij} \ne \emptyset$ for $\forall c_{ij} \ne \emptyset$; 2) for $\forall a \in P$, there exists $c_{ij} \ne \emptyset$ satisfying $(P - \{a\}) \bigcap c_{ij} = \emptyset$.

Theorem 1 implies that the core is the union of all singletons in the discernibility matrix, whereas Theorem 2 is a convenient way to test if a feature subset is a reduct. The first condition states that it is enough to use a reduct to together distinguish the sample pairs, of which the corresponding discernibility feature sets are not empty; the second one states that each feature in P is individually necessary. Therefore, a reduct is a minimal feature subset discerning these sample pairs, of which the corresponding discernibility feature sets are not empty.

IV. Relative Discernibility Relation Preserved Algorithm for FS-FRS

The relative discernibility relation of a feature was introduced in [13] and [57] to characterize indispensable and dispensable condition features. In this section, the relative discernibility relation is employed to characterize FS-FRS. By preserving the relative discernibility relation of the feature set, an algorithm is then developed to select a feature subset based on fuzzy rough sets.

Definition 2 ([57]): Let $(U, R \cup D)$ be a fuzzy decision table. The relative discernibility relation of $a \in R$ with respect to D is defined as $\mathrm{DIS}(a) = \{(x_i, x_j) : T(a(x_i, x_j), \lambda(x_i)) = 0, d(x_i) \neq d(x_j)\}$ where $\lambda(x_i) = Pos_R(D)(x_i)$.

The relative discernibility relation of a feature is actually the set of all sample pairs that can be discerned by this feature. That is, $(x_i, x_j) \in \mathrm{DIS}(a)$ means that the feature a can discern the sample pair (x_i, x_j). By Definition 2, $(x_i, x_j) \in \mathrm{DIS}(a)$ indicates that $a \in c_{ij}$ holds for every $c_{ij} \neq \emptyset$. This fact implies $c_{ij} = \{a \in R : (x_i, x_j) \in \mathrm{DIS}(a)\}$ and $\mathrm{DIS}(a) = \{(x_i, x_j) : a \in c_{ij}\}$. The relative discernibility relation of R is denoted by $\mathrm{DIS}(R) = \cup_{a \in R}\mathrm{DIS}(a)$. With the relative discernibility relation, we develop the following theorem to characterize the core.

Theorem 3: $\mathrm{Core}_D(R) = \{a \in R : \exists (x_i, x_j) \in \mathrm{DIS}(a), \text{s.t.}\ (x_i, x_j) \notin \mathrm{DIS}(R - \{a\})\}$.

By Theorem 3, if there exists a sample pair that can only be discerned by the feature a but cannot be discerned by any feature in $R - \{a\}$, then a belongs to the core. Based on the relative discernibility relation, we develop the following theorem to characterize FS-FRS.

Theorem 4: A subset $P \subseteq R$ is a reduct of R if and only if the following conditions hold: 1) $\mathrm{DIS}(P) = \mathrm{DIS}(R)$; 2) $\forall a \in P$, $\mathrm{DIS}(P - \{a\}) \neq \mathrm{DIS}(R)$.

Proof: 1) $\mathrm{DIS}(R) = \mathrm{DIS}(P) \Leftrightarrow \forall (x_i, x_j) \in \mathrm{DIS}(R)$, $(x_i, x_j) \in \mathrm{DIS}(P) \Leftrightarrow \forall (x_i, x_j) \in \mathrm{DIS}(R), \exists a \in P$, s.t., $(x_i, x_j) \in \mathrm{DIS}(a) \Leftrightarrow \forall c_{ij} \neq \emptyset, \exists a \in P$, s.t., $a \in c_{ij} \Leftrightarrow \forall c_{ij} \neq \emptyset, P \cap c_{ij} \neq \emptyset$.

2) $\forall a \in P$, $\mathrm{DIS}(P - \{a\}) \neq \mathrm{DIS}(R) \Leftrightarrow \forall a \in P$, $\exists (x_i, x_j) \in \mathrm{DIS}(R)$, s.t., $(x_i, x_j) \notin \mathrm{DIS}(P - \{a\}) \Leftrightarrow \forall a \in P, \exists c_{ij} \neq \emptyset$, s.t., $(P - \{a\}) \cap c_{ij} = \emptyset$. ■

According to Theorem 4, FS-FRS is a minimal feature subset discerning sample pairs that can be discerned by R. Based on this, we can design the relative discernibility relation preserved algorithm for finding one reduct with fuzzy rough sets.

By preserving the relative discernibility relation of R, Algorithm 1 computes one reduct from a fuzzy decision table. Step 1 computes the membership degree of each sample belonging to the positive region in a fuzzy decision table with a time complexity of $O(|U|^2\,|D|)$. Step 2 computes the relative discernibility relations of each feature and the feature set with a time complexity of $O(|U|^2\,|R|)$. Step 3 computes the core of the fuzzy decision table with a time complexity of $O(|U|^2\,|R|)$. Steps 5–8 always select a feature that maximizes the increment of sample pairs at each loop, with a time complexity of $O(|U|^2\,|R|)$. To sum up, the time complexity of Algorithm 1 is $O(|U|^2\,|R|)$.

Algorithm 1: Fuzzy Rough Set Based Feature Selection (FS-FRS).

Input: A fuzzy decision table $(U, R \cup D)$.
Output: A feature subset of $(U, R \cup D)$: red.

1: For each $x_i \in U$, compute $\lambda(x_i) = \underline{R_\vartheta}[x_i]_D(x_i)$;
2: For each feature $a \in R$, compute $DIS(a)$, and $DIS(R)$;
3: Compute $\mathrm{Core}_D(R)$ according to Theorem 3;
4: Let $red = \mathrm{Core}_D(R)$ and $DIS(red) = \bigcup_{a \in \mathrm{Core}_D(R)} DIS(a)$;
5: **while** $DIS(R) \neq DIS(red)$ **do**
6: Select $a_0 \in R - red$ satisfying
$$|DIS(red \cup \{a_0\})| = \max_{a \in R - \mathrm{red}}\{|DIS(red \cup \{a\})|\}$$
7: Let $red = red \cup \{a_0\}$, $DIS(red) = DIS(red) \cup DIS(a_0)$;
8: **end while**
9: **return** red.

Algorithm 1 is a novel structural method of discerning sample pairs for selecting features. Our experimental part will demonstrate its feasibility and effectiveness. Algorithm 1 further provides the theoretical foundation for the incremental perspective of FS-FRS in the next section, and will be a baseline algorithm in our experimental comparisons. As with all fuzzy rough set based batch methods, however, Algorithm 1 computes all samples of a dataset all at once. For large datasets, they are often costly, and even impractical since they easily exceed the memory capacity. In order to enhance the efficacy of the batch methods, the next section will study the incremental solution to FS-FRS.

V. Incremental Perspective for FS-FRS

This section investigates the incremental perspective for FS-FRS assuming data can be presented in sample subsets that arrive one after another. Rather than computing all samples of the dataset in batch, we handle sample subsets one by one so that FS-FRS can be performed from the incremental perspective. An incremental manner is first employed to compute the relative discernibility relations of each feature and the feature set with sample subsets arriving continuously. Based on the updated relations, an insight into the incremental process of feature selection is gained to reveal how to add and delete features. By the adoption of the incremental process, two incremental algorithms for FS-FRS are designed to employ the incremental fashion to compute a reduct from a real-valued dataset.

A. Incremental Environment and Notations

This section describes the incremental environment and the used symbols.

We assume that $(U, R \cup D)$ is a fuzzy decision table in this paper. In order to employ the incremental fashion to compute a reduct of $(U, R \cup D)$, we divide the universe U into a sample subset sequence $\{U_k\}_{k=1}^m$, each of which is called an incoming subset. Clearly, the following conditions must be satisfied:

1) $U = \cup_{k=1}^{m} U_k$;
2) $U_i \cap U_j = \emptyset$ for $i \neq j$.

In our incremental environment, sample subsets arrive one by one. The previously arriving subsets are stored in a temporary pool T that is initialized to an empty set. More specially, when U_1 is added into T, the temporary pool is updated as U_1; with U_2 arriving, the temporary pool is changed into $U_1 \cup U_2$; ...; with U_m arriving, the temporary pool is changed into the whole universe U. In a word, the temporary pool T always changes as the first, ..., mth subsets arrive successively.

In the temporary decision table $(T, R \cup D)$, $\lambda_T(x_i)$ is the membership degree of $x_i \in T$ belonging to the positive region of $(T, R \cup D)$, $\mathrm{DIS}_T(a)$ is the relative discernibility relation of $a \in R$, $\mathrm{DIS}_T(R)$ is the relative discernibility relation of R, and red_T is a reduct of $(T, R \cup D)$. Before adding any subset from U, all results from $(T, R \cup D)$ are empty since the temporary pool T is initialized to the empty set. In an incoming decision subtable $(U_k, R \cup D)$, $\lambda_k(x_i)$ is the membership degree of $x_i \in U_k$ belonging to the positive region of $(U_k, R \cup D)$, $\mathrm{DIS}_k(a)$ is the relative discernibility relation of $a \in R$, and $\mathrm{DIS}_k(R)$ is the relative discernibility relation of R. With U_k arriving, the temporary decision table is changed into $(T \cup U_k, R \cup D)$. Based on the previously results, we can incrementally compute the membership degree $\lambda(x_i)$ for $x_i \in T \cup U_k$, $\mathrm{DIS}(a)$, $\mathrm{DIS}(R)$, and red_T in $(T \cup U_k, R \cup D)$. It is evident that we obtain a reduct of $(U, R \cup D)$ when there is no incoming sample subset.

B. Updating the Relative Discernibility Relation

From what has been discussed in Section III, the membership degree of each sample belonging to the positive region is a necessary step in computing the relative discernibility relation of each feature. Thus, this section first incrementally computes the membership degree with the arrival of a sample subset, and then incrementally computes the relative discernibility relations of each feature and the feature set.

When a subtable $(U_k, R \cup D)$ is added into $(T, R \cup D)$, the following theorem gains an insight into the incremental computation of $\lambda(x_i)$ for $x_i \in T \cup U_k$.

Theorem 5: For $x_i \in T \cup U_k$, we have

$$\lambda(x_i) = \begin{cases} \lambda_T(x_i) \wedge \inf\limits_{u \in U_k, d(x_i) \neq d(u)} \vartheta(R(x_i, u), 0), & x_i \in T, \\ \lambda_k(x_i) \wedge \inf\limits_{u \in T, d(x_i) \neq d(u)} \vartheta(R(x_i, u), 0), & x_i \in U_k \end{cases}$$

where $a \wedge b$ is the minimum of a and b.

Proof: Since $\lambda_T(x_i) = \inf_{u \in T, d(u) \neq d(x_i)} \vartheta(R(x_i, u), 0)$ holds for $(T, R \cup D)$, for $\forall x_i \in T$ we have

$$\begin{aligned} \lambda(x_i) &= \inf_{u \notin [x_i]_D} \vartheta(R(x_i, u), 0) \\ &= \inf_{u \in T, d(u) \neq d(x_i)} \vartheta(R(x_i, u), 0) \wedge \\ &\quad\ \inf_{u \in U_k, d(u) \neq d(x_i)} \vartheta(R(x_i, u), 0) \\ &= \lambda_T(x_i) \wedge \inf_{u \in U_k, d(x_i) \neq d(u)} \vartheta(R(x_i, u), 0). \end{aligned}$$

Similarly, $\lambda(x_i) = \inf_{u \in U_k, d(u) \neq d(x_i)} \vartheta(R(x_i, u), 0)$ holds for $(U_k, R \cup D)$. Thus, for $\forall x_i \in U_k$, we have

$$\begin{aligned} \lambda(x_i) &= \inf_{u \notin [x_i]_D} \vartheta(R(x_i, u), 0) \\ &= \inf_{u \in T, d(u) \neq d(x_i)} \vartheta(R(x_i, u), 0) \wedge \\ &\quad\ \inf_{u \in U_k, d(u) \neq d(x_i)} \vartheta(R(x_i, u), 0) \\ &= \lambda_k(x_i) \wedge \inf_{u \in T, d(x_i) \neq d(u)} \vartheta(R(x_i, u), 0). \end{aligned}$$

Hence, we prove this theorem. ■

Theorem 5 implies $\lambda(x_i) \leq \lambda_T(x_i)$ holds for $\forall x_i \in T$, and $\lambda(x_i) \leq \lambda_k(x_i)$ holds for $\forall x_i \in U_k$. According to Theorem 5, in order to incrementally compute $\lambda(x_i)$ for $x_i \in T \cup U_k$, we only need to compute $\inf_{u \in T, d(x_i) \neq d(u)} \vartheta(R(x_i, u), 0)$ for $x_i \in U_k$ and $\inf_{u \in U_k, d(x_i) \neq d(u)} \vartheta(R(x_i, u), 0)$ for $x_i \in T$.

In what follows, we incrementally compute the relative discernibility relations of each feature and the feature set in $(T \cup U_k, R \cup D)$. The following theorem is given to show the relationship among $\mathrm{DIS}(a)$, $\mathrm{DIS}_k(a)$, and $\mathrm{DIS}_T(a)$.

Theorem 6: For $\forall a \in R$, we have $\mathrm{DIS}_T(a) \cup \mathrm{DIS}_k(a) \subseteq \mathrm{DIS}(a)$.

Proof: $\mathrm{DIS}_T(a) \cup \mathrm{DIS}_k(a) \subseteq \mathrm{DIS}(a)$ holds for $\forall a \in R$. For $\forall x_i \in T$, we have $\lambda(x_i) \leq \lambda_T(x_i)$. Thus, by the monotonity of T- norm, $T(a(x_i, x_j), \lambda(x_i)) = 0$ holds for $(T \cup U_k, R \cup D)$, which implies $(x_i, x_j) \in \mathrm{DIS}(a)$, i.e., $\mathrm{DIS}_T(a) \subseteq \mathrm{DIS}(a)$. In a similar way, for $\forall (x_i, x_j) \in \mathrm{DIS}_k(a)$, we have $T(a(x_i, x_j), \lambda_k(x_i)) = 0$, which implies $T(a(x_i, x_j), \lambda(x_i)) = 0$ since $\lambda(x_i) \leq \lambda_k(x_i)$ holds for $\forall x_i \in U_k$. Thus, we have $(x_i, x_j) \in \mathrm{DIS}(a)$, which implies $\mathrm{DIS}_k(a) \subseteq \mathrm{DIS}(a)$. Therefore, we can get $\mathrm{DIS}_T(a) \cup \mathrm{DIS}_k(a) \subseteq \mathrm{DIS}(a)$ for $\forall a \in R$. ■

By Theorem 6, for $\forall a \in R$, both $\mathrm{DIS}_T(a)$ and $\mathrm{DIS}_k(a)$ are the subsets of $\mathrm{DIS}(a)$. The following theorem incrementally computes the new sample pairs in $\mathrm{DIS}(a)$ for $\forall a \in R$.

Theorem 7: For $\forall a \in R$, the following statements hold:

1) for $x_i, x_j \in T$ and $d(x_i) \neq d(x_j)$, if $T(a(x_i, x_j), \lambda(x_i)) = 0$, then $(x_i, x_j) \in \mathrm{DIS}(a)$;

2) for $x_i, x_j \in U_k$ and $d(x_i) \neq d(x_j)$, if $T(a(x_i, x_j), \lambda(x_i)) = 0$, then $(x_i, x_j) \in \mathrm{DIS}(a)$;

3) for $x_i \in T, x_j \in U_k$, and $d(x_i) \neq d(x_j)$, if $T(a(x_i, x_j), \lambda(x_i)) = 0$, then $(x_i, x_j) \in \mathrm{DIS}(a)$; if $T(a(x_j, x_i), \lambda(x_j)) = 0$, then $(x_j, x_i) \in DIS(a)$.

Proof: By Definition 2, we can easily prove this theorem. ■

By Theorem 7, we design Algorithm 2 to update the relative discernibility relations of each feature and the feature set.

Algorithm 2 is mainly made up of two parts: 1) updating $\lambda(x_i)$ for $\forall x_i \in T \cup U_k$; 2) updating $\mathrm{DIS}(a)$ for $\forall a \in R$, and $\mathrm{DIS}(R)$. The time complexity of part 1) is $O(|T||U_k||R|)$, and that of part 2) is $O((|T|^2 - |\mathrm{DIS}_T(a)|)|R|)$. Therefore, the time complexity of Algorithm 2 is $\max(O(|T||U_k||R|), O((|T|^2 - |\mathrm{DIS}_T(a)|)|R|))$.

C. Incremental Process of FS-FRS

When a sample subset U_k enters into the temporary pool T, the current reduct red_T either satisfies $\mathrm{DIS}(R) = \mathrm{DIS}(\mathrm{red}_T)$, or

Algorithm 2: Updating the Relative Discernibility Relation.

Input: 1) A current sample set T; 2) $\lambda_T(x_i)$ for each $x_i \in T$; 3) $DIS_T(a)$ for each $a \in R$, and $DIS_T(R)$; 4) An incoming sample subset U_k.
Output: $DIS(a)$ for $a \in R$, and $DIS(R)$.
1: According to Definition 2, compute $DIS_k(a)$ for $\forall a \in R$, and $DIS_k(R)$ in $(U_k, R \cup D)$;
2: By Theorem 6, compute $\lambda(x_i)$ for each $x_i \in T \cup U_k$;
3: Let $DIS(R) = DIS_T(R) \cup DIS_k(R)$;
4: **for** each $a \in R$ **do**
5: Let $DIS(a) = DIS_T(a) \cup DIS_k(a)$;
6: For each $(x_i, x_j) \notin DIS_T(a)$ satisfying $\forall x_i, x_j \in T$ and $d(x_i) \neq d(x_j)$, if $T(a(x_i, x_j), \lambda(x_i)) = 0$, then let $DIS(a) = DIS(a) \cup (x_i, x_j)$ and $DIS(R) = DIS(R) \cup (x_i, x_j)$;
7: For each $(x_i, x_j) \notin DIS_k(a)$ satisfying $\forall x_i, x_j \in U_k$ and $d(x_i) \neq d(x_j)$, if $T(a(x_i, x_j), \lambda(x_i)) = 0$, then let $DIS(a) = DIS(a) \cup (x_i, x_j)$ and $DIS(R) = DIS(R) \cup (x_i, x_j)$;
8: For $x_i \in T, x_j \in U_k$, and $d(x_i) \neq d(x_j)$, if $T(a(x_i, x_j), \lambda(x_i)) = 0$, then let $DIS(a) = DIS(a) \cup (x_i, x_j)$ and $DIS(R) = DIS(R) \cup (x_i, x_j)$; if $T(a(x_j, x_i), \lambda(x_j)) = 0$, then let $DIS(a) = DIS(a) \cup (x_j, x_i)$ and $DIS(R) = DIS(R) \cup (x_j, x_i)$;
9: **end for**
10: **return** $DIS(a)$ for each $a \in R$, and $DIS(R)$.

not. That is, there are two cases: 1) $\mathrm{DIS}(R) = \mathrm{DIS}(\mathrm{red}_T)$; 2) $\mathrm{DIS}(R) \neq \mathrm{DIS}(\mathrm{red}_T)$. For the two cases, we study the incremental process of feature selection by developing the strategies of adding and deleting features.

Case 1: $\mathrm{DIS}(R) = \mathrm{DIS}(\mathrm{red}_T)$.

Theorem 8: $\mathrm{DIS}(R) = \mathrm{DIS}(\mathrm{red}_T)$ implies that red_T contains a reduct of $(T \cup U_k, R \cup D)$.

Proof: $\mathrm{DIS}(R) = \mathrm{DIS}(\mathrm{red}_T)$ indicates any sample pair that can be discerned by R can also be discerned by red_T. By Theorem 4, red_T contains a reduct of $(T \cup U_k, R \cup D)$.

According to Theorem 8, red_T is either a reduct of the updated temporary decision table $(T \cup U_k, R \cup D)$, or it properly contains a reduct of $(T \cup U_k, R \cup D)$. To compute a reduct of $(T \cup U_k, R \cup D)$, we develop the following strategy.

First Strategy of Feature Deletion: Feature $a \in \mathrm{red}_T$ can be deleted from red_T if $\mathrm{DIS}(\mathrm{red}_T - \{a\}) = \mathrm{DIS}(R)$.

The First Strategy of Feature Deletion indicates that $\mathrm{red}_T - \{a\}$ can still discern all sample pairs that are discerned by R in the updated temporary decision table, so that the feature $a \in \mathrm{red}_T$ can be deleted from red_T. If the strategy does not hold for $\forall a \in \mathrm{red}_T$, then by Theorem 4, red_T is also a reduct of $(T \cup U_k, R \cup D)$. Otherwise, if the strategy holds for $\exists a \in \mathrm{red}_T$, then any sample pair that is discerned by R can also be discerned by $\mathrm{red}_T - \{a\}$, which implies that a can be deleted from red_T. We can continue the strategy of deleting features until the strategy does not hold. Thus, we can obtain a reduct of the updated temporary decision table.

Case 2: $\mathrm{DIS}(R) \neq \mathrm{DIS}(\mathrm{red}_T)$.

Case 2 implies red_T is not a reduct of the updated temporary decision table, since it does not satisfy condition 1) of Theorem 4. Thus, we need to add some features $B \subseteq R - \mathrm{red}_T$ into red_T until $\mathrm{DIS}(R) = \mathrm{DIS}(\mathrm{red}_T \cup B)$. By Theorem 4, we develop the following strategy of adding features.

Strategy of Feature Addition: Feature subset $B \in R - \mathrm{red}_T$ can be added into red_T if B is a minimal addition subset satisfying $\mathrm{DIS}(\mathrm{red}_T \cup B) = \mathrm{DIS}(R)$.

By Strategy of Feature Addition, $\mathrm{red}_T \cup B$ can discern all sample pairs that are discerned by R. This fact implies $\mathrm{red}_T \cup B$ contains a reduct of $(T \cup U_k, R \cup D)$. That is, there may be redundant features in $\mathrm{red}_T \cup B$. The following strategy is developed to obtain a reduct of $(T \cup U_k, R \cup D)$.

Second Strategy of Feature Deletion: Feature a can be deleted from $\mathrm{red}_T \cup B$ if $\mathrm{DIS}((\mathrm{red}_T \cup B) - \{a\}) = \mathrm{DIS}(R)$.

If the strategy of deleting features does not hold for $\forall a \in \mathrm{red}_T \cup B$, then $\mathrm{red}_T \cup B$ is a reduct of the updated temporary decision table by Theorem 4. If the strategy holds for $\exists a \in \mathrm{red}_T \cup B$, then a should be deleted from $\mathrm{red}_T \cup B$ by Theorem 4. By continuing Second Strategy of Feature Deletion, we can obtain a reduct of the updated temporary decision table.

To be brief, if an incoming subset is in Case 1, we apply First Strategy of Feature Deletion to obtain a reduct of the updated temporary decision table; if an incoming subset is in Case 2, we can first apply Strategy of Feature Addition to add the feature set $B \in R - \mathrm{red}_T$, and then use Second Strategy of Feature Deletion to delete redundant features from $\mathrm{red}_T \cup B$.

Remark 1: Removing redundant features is the key task for selecting an optimal feature subset [62]. Due to the measurement error, the complete removal of redundancy is not good. Hence, it is suggested that a useful principle is needed to control the level of redundancy in the process of selecting features. In fact, there may be more than one optimal feature subset on a dataset. In the framework of fuzzy rough sets, the redundancy is related to a selected feature subset [63]. That is, a feature is redundant related to an optimal feature subset, but not redundant related to other optimal feature subsets. In the proposed methods, a feature a is said to be *redundant* related to a feature subset P, if $\mathrm{DIS}(P) = \mathrm{DIS}(R)$ and $\mathrm{DIS}(P) = \mathrm{DIS}(P \cup \{a\})$. The definition of redundancy implies that the sample pairs discerned by the redundant feature can always be discerned by an optimal feature subset. Therefore, our proposed methods can control the level of redundancy, and remove the redundant features related to a selected feature subset. Features can be categorized into essential feature, derogatory feature, indifferent feature, and redundant feature. In the framework of fuzzy rough set, the essential feature naturally belongs to the core that is obtained according to Theorem 3. Our proposed methods can discard derogatory feature by judging whether sample pairs discerned by it can be discerned by any feature subset. Indifferent feature can be discarded since it does not provide the discernibility information about sample pairs in the proposed framework.

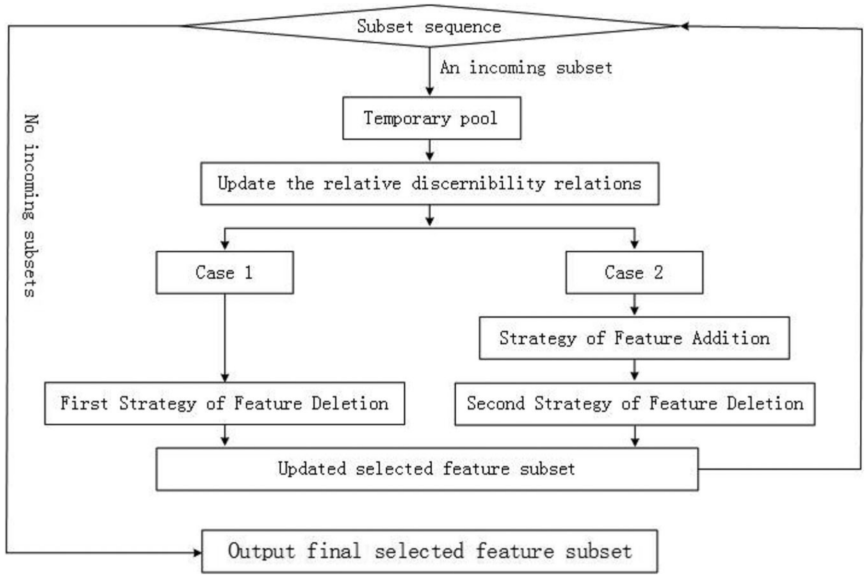

Fig. 1. First incremental version of FS-FRS.

D. Incremental Perspective for FS-FRS

This section investigates FS-FRS in the incremental environment. Two incremental versions of FS-FRS are designed by judging whether to compute a reduct with sample subsets arriving sequentially. In terms of the two incremental versions, we present two corresponding algorithms to incrementally select a feature subset from a dataset.

Version 1: Perform feature selection with each sample subset arriving.

In this version, we update the selected feature subset each time a sample subset arrives. That is, with the arrival of each sample subset, the reduct of a temporary decision table is updated based on the incremental computation of the relative discernibility relations. If it is Case 1, we perform First Strategy of Feature deletion; if it is Case 2, we perform Strategy of Feature Addition and then Second Strategy of Feature Deletion. This process is repeated until there is no incoming subset left, resulting in a final reduct from the whole dataset. The flow chart of Version 1 is shown in Fig. 1.

By incremental version, we propose the following incremental algorithm for FS-FRS.

Algorithm 3 computes a reduct starting from an empty temporary pool. Step 1 initializes T and these results from $(T, R \cup D)$ as empty sets. Step 3 selects a sample subset that will be added into T. Step 4 incrementally computes the relative discernibility relations of each feature and the feature set in the updated temporary decision table, with a time complexity of $\max(O(|T||U_k||R|), O((|T|^2 - |\mathrm{DIS}_T(a)|)|R|))$. Step 6 decides whether it is Case 1 or Case 2, with the time complexity of $O(|(T \cup U_k)^2||R|)$. Steps 6–10 perform the strategies of adding features, with a time complexity of $O(|R|)$. Steps 11–14 perform the strategies of deleting features, with a time complexity of $O(|R|)$. Hence, the time complexity of Algorithm 3 is $\max(O(|T||U_k||R|), O((|T|^2 - |\mathrm{DIS}_T(a)|)|R|))$.

At each iteration, however, Version 1 updates a selected feature subset, which is often time-consuming. To accelerate

Algorithm 3: First Incremental Version for Fuzzy Rough Set Based Feature Selection (IV-FS-FRS-1).

Input: 1) The sample subset sequence: $U = \{U_k\}_{k=1}^{m}$; 2) The feature set: R; 3) The decision set: D;
Output:
A reduct of $(U, R \cup D)$: red;
1: Initialize: 1) $red = \emptyset$; 2) The pool to store the previously incoming samples $T = \emptyset$; 3) $DIS_T(a) = \emptyset$ for each $a \in R$, and $DIS_T(R) = \emptyset$; 4) Iterations $k = 1$;
2: **while** U is not empty **do**
3: Observe a sample subset U_k from U, and add it into T;
4: According to Algorithm 2, compute $DIS(a)$ for each $a \in R$ and $DIS(R)$ in $(T \cup U_k, R \cup D)$;
5: Let $DIS_T(a) = DIS(a)$ for each $a \in R$, and $DIS_T(R) = DIS(R)$;
6: **while** $DIS_T(R) \neq DIS_T(red)$ **do**
7: For each $a \in R - red$, compute $DIS_T(red \cup \{a\})$;
8: Select the feature $a_0 \in R - red$ satisfying
$$|DIS_T(red \cup \{a_0\})| = \max_{a \in R - \mathrm{red}} |DIS_T(red) \cup DIS_T(a)|;$$
9: Let $red = red \cup \{a_0\}$, and $DIS_T(red) = DIS_T(red) \cup DIS_T(a_0)$;
10: **end while**
11: **while** $DIS_T(R) = DIS_T(red)$ **do**
12: For each $a \in red$, compute $DIS_T(red - \{a\})$;
13: Select $a_0 \in red$ satisfying $DIS_T(R) = DIS_T(red - \{a_0\})$, and let $red = red - \{a_0\}$;
14: **end while**
15: Let $U = U - U_k$, $T = T \cup U_k$, $k = k + 1$;
16: **end while**
17: **return** red.

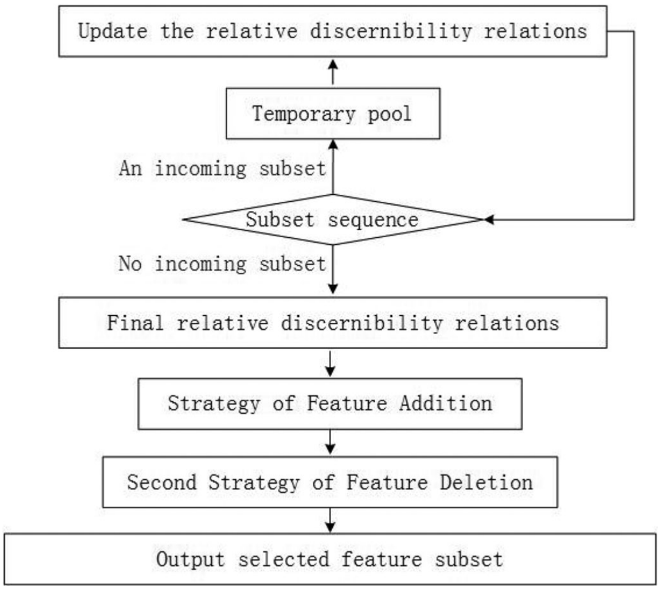

Fig. 2. Second incremental version of FS-FRS.

Version 1, the feature subset is not updated each time a sample subset comes, yielding the following incremental version.

Version 2: Perform feature selection when there is no further sample subset arriving.

This version begins with an empty set to compute the selected features from the whole dataset when no incoming subset is left. Upon the arrival of each subset, we only update the relative discernibility relations of each feature and the feature set. When there is no incoming subset, we obtain the final relative discernibility relations of each feature and the feature set on the whole dataset. Based on the final relations, we utilize Strategy of Feature Addition and Second Strategy of Feature Deletion to compute a final reduct of the fuzzy decision table. The flow chart is shown in Fig. 2.

The following algorithm is formulated to compute a reduct from a dataset based on Version 2.

In Algorithm 4, Step 1 initializes T and these results from $(T, R \cup D)$ as empty sets. Steps 2–4 updates the relative discernibility relations of each feature and the feature set, with the time complexity of $\max(O(|T||U_k||R|), O((|T|^2 - |\mathrm{DIS}_T(a)|)|R|))$. Steps 8–12 perform Strategy of Feature Addition, with the time complexity of $O(|R|)$. Steps 13–16 perform Second Strategy of Feature Deletion, with a time complexity of $O(|R|)$. Hence, the time complexity of Algorithm 4 is $\max(O(|T||U_k||R|), O((|T|^2 - |\mathrm{DIS}_T(a)|)|R|))$.

Remark 2: Two proposed incremental algorithms expedite FS-FRS, which will be demonstrated in the next section. However, there are different operation processes between them. Upon the arrival of a sample subset, Version 1 updates the selected feature subset, whereas Version 2 only performs feature selection where there is no further sample subset. Obviously, Version 2 saves some runtime without updating the feature subset each time a sample subset arrives, which can be shown in our experimental comparisons. In fact, Version 1 is more applicable to dynamic datasets with sample arriving, since it can find a

Algorithm 4: Second Incremental Version for Fuzzy Rough Set Based Feature Selection (IV-FS-FRS-2).

Input: 1) The sample set sequence $U = \{U_k\}_{k=1}^{m}$; 2) The feature set: R; 3) The decision set D;
Output: A reduct of $(U, R \cup D)$: red;
1: Initialize: 1) $red = \emptyset$; 2) The temporary pool to store the previously incoming samples $T = \emptyset$; 3) $DIS_T(a) = \emptyset$ for each $a \in R$, and $DIS(R) = \emptyset$; 4) Iterations $k = 1$;
2: **while** U is not empty **do**
3: Observe a sample subset U_k from U, and add it into T;
4: By Algorithm 2, compute $DIS(a)$ for each $a \in R$ and $DIS(R)$ in $(T \cup U_k, R \cup D)$;
5: Let $DIS_T(a) = DIS(a)$ for each $a \in R$, and $DIS_T(R) = DIS(R)$;
6: Let $U = U - U_k$, $T = T \cup U_k$, and $k = k + 1$;
7: **end while**
8: **while** $DIS_T(R) \neq DIS_T(red)$ **do**
9: For each $a \in R - red$, compute $DIS_T(red \cup \{a\})$;
10: Select the feature $a_0 \in R - red$ satisfying

$$|DIS_T(red \cup \{a_0\})| = \max_{a \in R - \mathrm{red}} |DIS_T(red) \cup DIS_T(a)|;$$

11: Let $red = red \cup \{a_0\}$, and $DIS_T(red) = DIS_T(red) \cup DIS_T(a_0)$;
12: **end while**
13: **while** $DIS_T(R) = DIS_T(red)$ **do**
14: For each $a \in red$, compute $DIS_T(red - \{a\})$;
15: Select $a_0 \in red$ satisfying $DIS_T(R) = DIS_T(red - \{a_0\})$, and let $red = red - \{a_0\}$;
16: **end while**
17: **return** red.

real-time feature subset. Unlike Version 1, Version 2 is more suitable to large datasets that can be presented in sample subsets to be added sequentially, with selecting the feature subset where there is no further sample subset arriving.

Remark 3: It is a good idea to introduce the true sliding window into the incremental operation of FS-FRS, in which we can use the proposed strategies of adding and deleting features to update the selected feature subset within a window. However, a temporally localized reduct that is often different from a real reduct of the dataset is obtained based on samples within this current window. The quality of the temporary reduct is expected to be low. This implies that the choice of the window size is not arbitrary. It usually depends on the domain in question and the samples of all data streams, which is clearly nontrivial. Once the window size n_w is determined, the sliding window operation commences so that at most n_w samples are visible to the incremental computation of feature selection. During the sliding window, however, it is problematic in determining which outdated samples are filtered out and which future samples are received. From this perspective, it is time-consuming in performing the sliding window operation, since one has to consider the relation between each sample and all samples of the whole dataset in the framework of fuzzy rough sets. Therefore, to solve the above shortcomings, we will study the combination of the incremental operation and the sliding window in the future work.

VI. Experimental Evaluation

In this section, we experimentally evaluate the performance of our proposed incremental algorithms for FS-FRS on several UCI datasets. Four batch algorithms are selected as baseline algorithms for comparisons with second incremental version for FS-FRS (IV-FS-FRS-2) to show the effectiveness of the proposed incremental version. They are FS-FRS (Algorithm 1), the discernibility matrix based algorithm [55] (denoted by Matrix), the dependence function based algorithm [58] (denoted by Dependence), and fuzzy entropy based algorithm [61] (denoted by Entropy) that is actually stochastic. The efficiency of IV-FS-FRS-2 is further demonstrated via comparisons with first incremental version for FS-FRS (IV-FS-FRS-1) (Algorithm 3) and the neighbor rough set based incremental algorithm [60] (denoted by INFS) that incrementally selects a feature subset from a real-valued dataset with samples arriving one by one.

A. Experimental Design

The hardware environment: Intel (R) Xeon (R) CPU X5690 @3.47 GHz 3.46 GHz (2 processors), and 64.0 GB.

The software environment: MATLAB 2012b.

Dataset: Ten real-valued datasets from UCI Machine Learning Repository are used (see Table I) in our experimental comparisons.

Data split: To perform our proposed IV-FS-FRS-1 and IV-FS-FRS-2, each dataset in Table I is randomly divided into 10 sample subsets with equal size. So, these subsets are made up of a sample subset sequence.

Classifier: K-nearest neighbor classifier (K is set as 3) is used to test the classification accuracy of features selected by several comparison algorithms. Tenfold cross validation is applied to the ten datasets.

Fuzziness of the real-valued dataset: For a real-valued feature a, the feature value of each sample is normalized as $\overline{a}(x_i) = \frac{a(x_i)-\min_j a(x_j)}{\max_j a(x_j)-\min_j a(x_j)}, x_i \in U$. Thus, $\overline{a}(x_i) \in [0,1]$ for $\forall x_i \in U$. A fuzzy similarity relation for $\overline{a}$ is defined as $\overline{a}(x_i, x_j) = 1 - |\overline{a}(x_i) - \overline{a}(x_j)|$ for $x_i, x_j \in U$.

Triangular norm: T_L is used in our experiments.

B. Comparisons of IV-FS-FRS-2 and Four Batch Algorithms

This section evaluates the effectiveness of IV-FS-FRS-2 on ten datasets in terms of the following aspects. One is to compare the runtime of IV-FS-FRS-2 and four batch algorithms. The other is to compare the size and the accuracy of the features selected by IV-FS-FRS-2, FS-FRS, Matrix, Dependence, and Entropy. The experimental results are summarized in Tables II–IV, where the symbol "*" means this method cannot select a feature subset from the dataset in the current software and hardware environments.

TABLE I
Description of the Datasets

Dataset	Abbreviation	Samples	Feature	Class
Sonar	Sonar	208	60	2
Spect	Spect	267	44	2
Iono	Iono	351	34	2
Libras movement	Libras	360	90	15
WDBC	WDBC	569	30	2
QSAR biodegradation	QSAR	1055	41	2
Parkinsons multiple	Park	1208	28	2
Waveform	Wave	5000	21	3
Thyroid	Thyroid	9172	28	50
Gamma	Gamma	19 020	10	2

TABLE II
Runtime (S) of IV-FS-FRS-2 and Four Batch Algorithms

Dataset	IV-FS-FRS-2	FS-FRS	Matrix	Dependence	Entropy
Sonar	3.08	66.40	15.42	682.08	8.47
Spect	2.01	75.58	21.22	480.70	11.03
Iono	2.07	77.71	27.42	515.49	14.03
Libras	15.35	439.56	216.95	743.32	20.59
WDBC	7.12	285.36	73.97	120.14	37.53
QSAR	28.23	136.41	516.37	787.97	311.16
Park	21.67	91.06	467.23	336.44	311.70
Wave	199.88	6233.82	3121.04	4365.85	5779.02
Thyroid	15 705.43	66 795.98	43 878.87	44 923.65	83 400.09
Gamma	19 174.40	*	*	*	*

Table II presents that in comparison with the four batch algorithms, IV-FS-FRS-2 greatly reduces the runtime of selecting features from each selected dataset. For example, the runtime of IV-FS-FRS-2 shows a decrease up to 3.49%, 7.08%, 2.07%, and 74.55% of that FS-FRS, Matrix, Dependence, and Entropy on "Libras," respectively; the runtime of IV-FS-FRS-2 decreases to 20.69%, 5.47%, 3.58%, and 9.07% of that of FS-FRS, Matrix, Dependence, and Entropy on "QSAR," respectively. Moreover, we can see from Table II that on "Gamma," FS-FRS, Matrix, Dependence, and Entropy cannot obtain a selected feature subset, but IV-FS-FRS-2 can compute a feature subset under the current software and hardware environments. This fact implies that it is possible to employ IV-FS-FRS-2 to obtain a selected feature subset from large-scale datasets. Therefore, IV-FS-FRS-2 can not only expedite feature selection based on fuzzy rough sets, but also deal with large-scale datasets.

Table III presents that the average size of features obtained by IV-FS-FRS-2 (23.8) is less than that of feature subsets selected by "Raw" (38.6), FS-FRS (37.11), Matrix (37.67), Dependence (31.89), and Entropy (27.33). The reason is that IV-FS-FRS-2 has the step of removing redundant features. Furthermore, Table IV presents the accuracy of features selected by IV-FS-FRS-2, FS-FRS, Matrix, Dependence, and Entropy, where "Raw" represents the accuracy of all features and "Advantages" is the number of times that the accuracy of a method is superior to "Raw" on the ten datasets. From Table IV, we can see that the accuracy of IV-FS-FRS-2 is higher than "Raw" on five selected datasets; the accuracy of FS-FRS is better than "Raw" on five datasets; both Matrix and Dependence have advantages

TABLE III
SIZE OF FEATURES SELECTED BY IV-FS-FRS-2 AND FOUR BATCH ALGORITHMS

Dataset	Raw	IV-FS-FRS-2	FS-FRS	Matrix	Dependence	Entropy
Sonar	60	20	57	57	52	19
Spect	44	31	44	44	38	30
Iono	34	27	28	29	30	25
Libras	90	51	72	72	47	52
WDBC	30	24	30	30	25	24
QSAR	41	34	40	35	33	34
Park	28	24	27	27	17	24
Wave	21	17	19	18	17	21
Thyroid	28	5	17	27	28	17
Gamma	10	5	*	*	*	*
Average	38.6	23.8	37.11	37.67	31.89	27.33

TABLE IV
CLASSIFICATION ACCURACY OF FEATURES SELECTED BY IV-FS-FRS-2 AND FOUR BATCH ALGORITHMS

Dataset	Raw	IV-FS-FRS-2	FS-FRS	Matrix	Dependence	Entropy
Sonar	0.8612	0.7983	0.8364	0.8417	0.8324	0.7976
Spect	0.7120	0.7376	0.7148	0.7303	0.7151	0.7150
Iono	0.6379	0.8545	0.8574	0.8660	0.8547	0.6406
Libras	0.7944	0.7639	0.7722	0.7806	0.7806	0.7611
WDBC	0.9596	0.9666	0.9684	0.9737	0.9701	0.9649
QSAR	0.8455	0.8427	0.8371	0.8388	0.8379	0.8417
Park	0.9835	0.9859	0.9851	0.9843	0.9876	0.9653
Wave	0.7960	0.7972	0.8008	0.7614	0.7738	0.7946
Thyroid	0.8210	0.8209	0.7606	0.8201	0.8204	0.7813
Gamma	0.8302	0.7985	*	*	*	*
Advantages		5	5	4	4	3

over "Raw" on four datasets; and Entropy outperforms "Raw" on three datasets. These facts indicate that IV-FS-FRS-2 and FS-FRS can improve the accuracy of "Raw," with ignoring the fact that Matrix and Dependence are better than IV-FS-FRS-2 and FS-FRS on "Iono," "WDBC," and "Park." Furthermore, on "Sonar," "Libras," "QSAR," and "Thyroid," all of several comparisons algorithms are close to "Raw," where IV-FS-FRS-2 and FS-FRS are superior to Matrix, Dependence, and Entropy on "QSAR" and "Thyroid." Hence, we can draw the conclusion from the above facts that our proposed methods can improve the accuracy of "Raw" on most of the selected datasets.

C. Comparison of IV-FS-FRS-2, IV-FS-FRS-1, and INFS

In this section, IV-FS-FRS-2 is compared with IV-FS-FRS-1 and INFS on the ten datasets. These comparisons focus on the runtime, the size and accuracy of the selected features, as well as the discussion that is suggested to show how the sample subset sequence size has an influence on the overall efficacy of the proposed incremental algorithms. The experimental results are summarized in Table V, and Figs. 3–5.

Table V presents that IV-FS-FRS-2 is more efficient than IV-FS-FRS-1, and IV-FS-FRS-1 is more efficient than INFS on the ten datasets. There may be the following reasons. The first one is with each subset arriving, IV-FS-FRS-2 only updates the relative discernibility relations without performing feature selection, whereas IV-FS-FRS-1 updates the selected feature subset by updating the relative discernibility relations. The second one is upon the arrival of a sample subset, our proposed incremental algorithms can handle it all at once, while INFS needs to deal with them one at a time. The third one is our proposed incremental algorithms are designed based on the strategies of adding and deleting features, whereas INFS has no investigations for how to add and delete features. Hence, in comparisons with INFS, our proposed incremental algorithms can save the runtime of obtaining a feature subset from a real-valued dataset, and IV-FS-FRS-2 is much faster than IV-FS-FRS-1 and INFS.

Moreover, Table V presents different sizes of features selected by three incremental algorithms, since with a sample subset arriving IV-FS-FRS-1 and INFS updates the selected feature subset while IV-FS-FRS-2 performs feature selection where no subset is left. On most of the selected datasets, the accuracy of our proposed incremental algorithms is higher than that of INFS. Therefore, IV-FS-FRS-2 can find a comparable feature subset from a real-valued dataset in a much shorter time.

Fig. 3 depicts the detailed changes of the runtime of IV-FS-FRS-1 and IV-FS-FRS-2 with subsets continuously arriving. Fig. 3 shows IV-FS-FRS-2 is faster than IV-FS-FRS-1 with the first nine subsets arriving. The reason is with the arrival of the first nine subsets, IV-FS-FRS-1 updates selected features based on the incremental computation of the relative discernibility relations, whereas IV-FS-FRS-2 only incrementally computes the relative discernibility relations without performing feature selection. Thereby, IV-FS-FRS-2 saves the runtime with the first nine subsets arriving. Furthermore, we can also observe that compared with IV-FS-FRS-1, IV-FS-FRS-2 consumes more runtime with the arrival of the final subset of each dataset, since IV-FS-FRS-2 performs feature selection starting with an empty set, whereas IV-FS-FRS-1 updates the selected features based on a current feature subset.

To show the influence of the sequence size on the efficacy of our incremental algorithms, each dataset is randomly divided into 2–10 parts with equal size. The results are depicted in Figs. 4 and 5, where the x-axis is the sequence size. Fig. 4 displays IV-FS-FRS-1 and IV-FS-FRS-2 are consistently much faster than INFS with changing the sequence size on each dataset. There may be the following reasons. One is with a sample subset arriving, our proposed incremental algorithms can handle them all at once, whereas INFS has to be run for many times. The other is our proposed incremental algorithms are designed by the strategies of adding and deleting features, whereas INFS has no investigation for how to add and delete features. Furthermore, Fig. 4 indicates that the runtime of IV-FS-FRS-1 and IV-FS-FRS-2 basically increases with increasing the sequence size. From Fig. 4, we can also conclude that our proposed algorithms are more efficient than INFS when the sequence size ranges from 2 to 4.

Fig. 5 shows the number and accuracy of features selected by IV-FS-FRS-1, IV-FS-FRS-2, and INFS as the sequence size changes from 2 to 10. Fig. 5 indicates the number of features selected by IV-FS-FRS-2 and INFS is basically stable with changing the sequence size, whereas the number of features selected by IV-FS-FRS-1 fluctuates significantly. Moreover, the accuracy of features selected by IV-FS-FRS-2 and INFS changes a little,

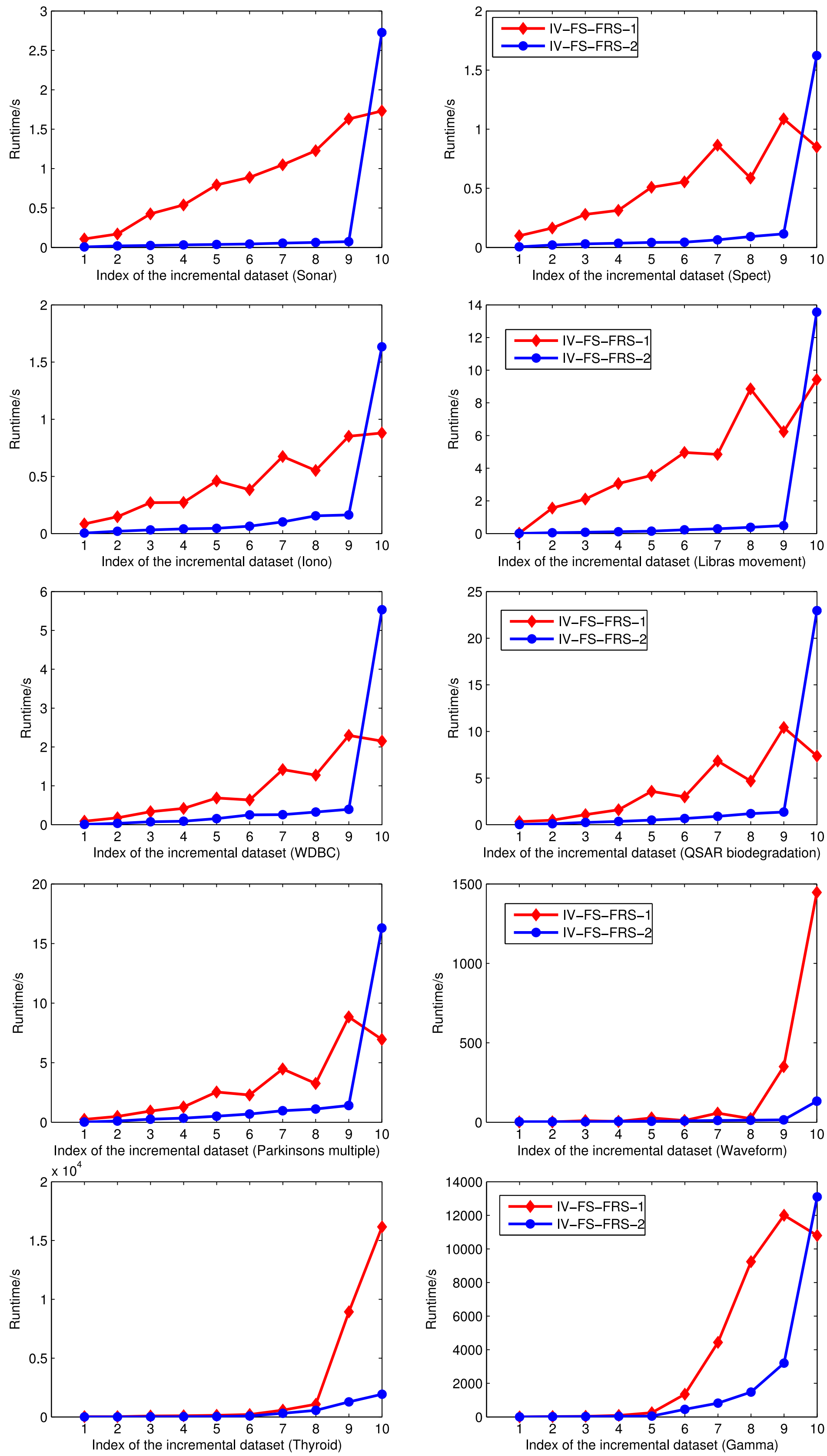

Fig. 3. Runtime of IV-FS-FRS-1 and IV-FS-FRS-2 with respect to subsets continuously arriving.

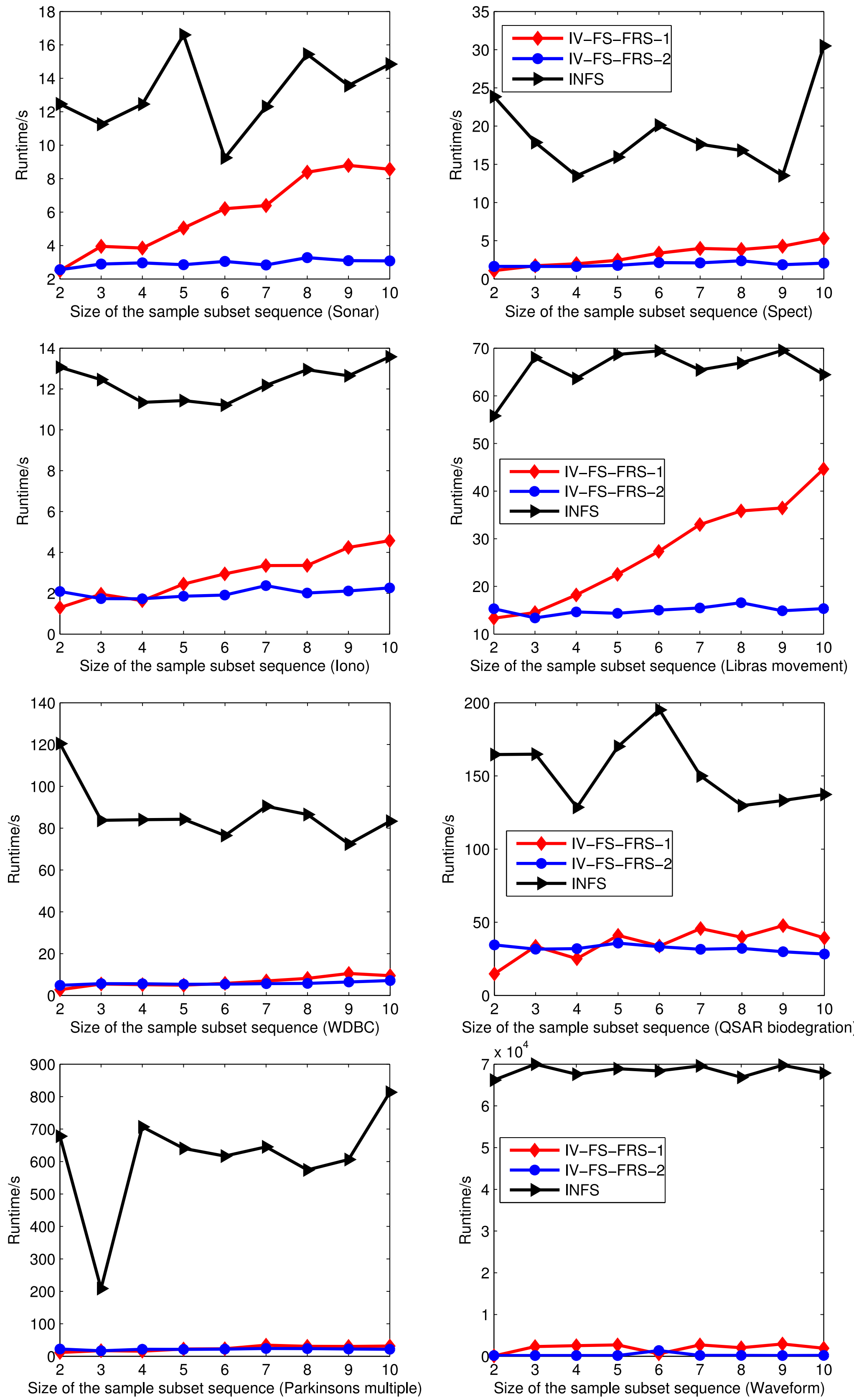

Fig. 4. Runtime of IV-FS-FRS-1, IV-FS-FRS-2, and INFS as the size of the sample subset sequence changes.

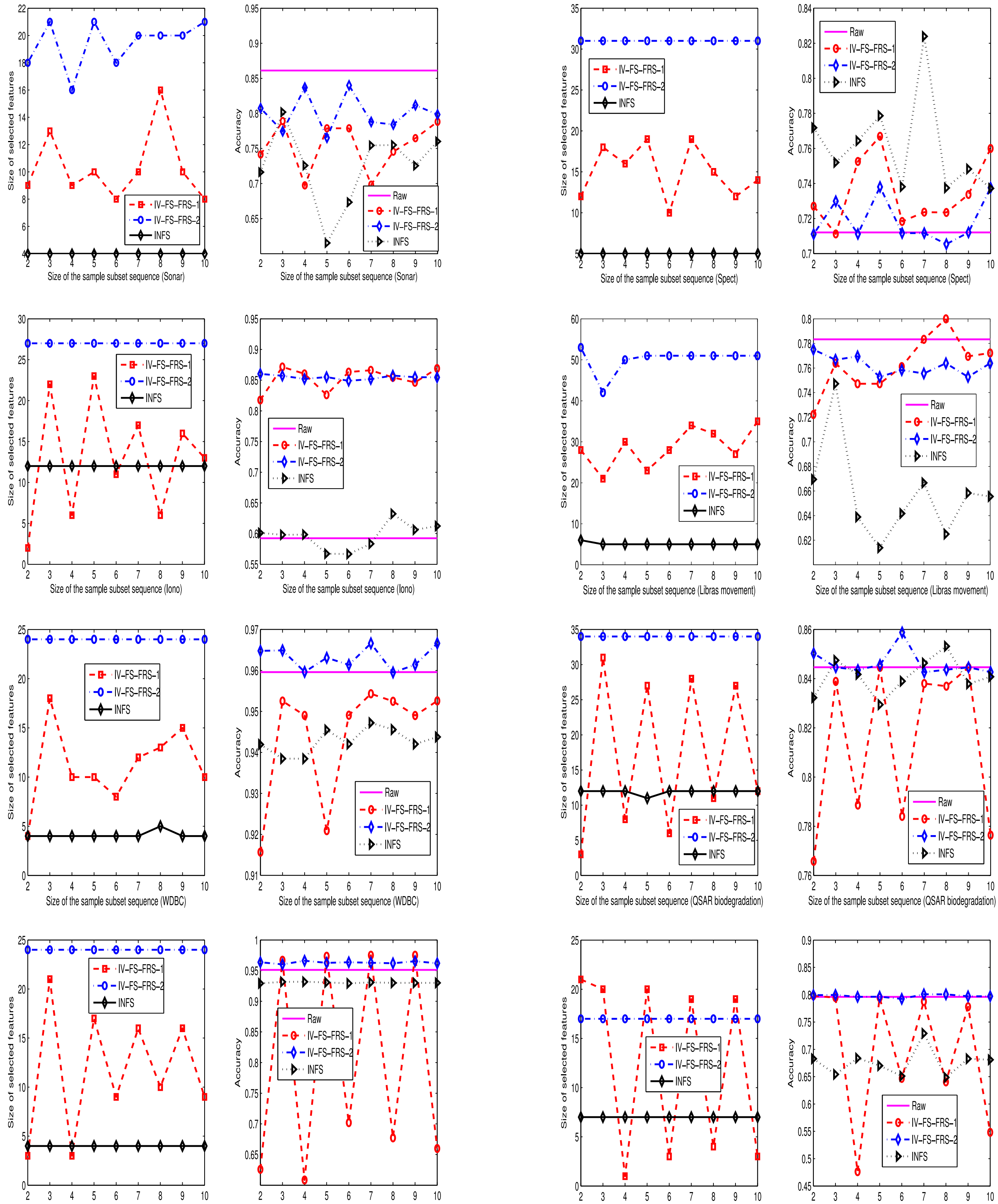

Fig. 5. Number and accuracy of features selected by IV-FS-FRS-1, IV-FS-FRS-2, and INFS as the size of the sample subset sequence changes.

TABLE V
EXPERIMENTAL RESULTS OF IV-FS-FRS-1, IV-FS-FRS-2, AND INFS

Dataset	IV-FS-FRS-1			IV-FS-FRS-2			INFS		
Runtime/s	No.	Accuracy	Runtime/s	No.	Accuracy	Runtime/s	No.	Accuracy	
Sonar	8.56	8	0.7881	3.08	20	0.7983	14.85	4	0.76
Spect	5.31	14	0.7600	2.01	31	0.7376	30.50	5	0.7372
Iono	4.57	13	0.8690	2.07	27	0.8545	13.58	30	0.6123
Libras	44.64	35	0.7722	15.35	51	0.7639	64.44	5	0.6556
WDBC	9.48	15	0.9526	7.12	24	0.9666	83.35	4	0.9438
QSAR	39.30	12	0.7764	28.23	34	0.8427	137.30	12	0.8408
Park	31.25	9	0.6598	21.67	24	0.9859	813.37	4	0.9301
Wave	1930.77	3	0.5482	199.88	17	0.7972	67 861.76	7	0.6810
Thyroid	29 291.57	4	0.8205	15705.43	5	0.8209	84 539.64	23	0.8132
Gamma	38 238.69	6	0.8315	19174.40	5	0.7985	65 438.97	10	0.8364

whereas the accuracy of IV-FS-FRS-1 fluctuates significantly. The accuracy of IV-FS-FRS-2 outperforms that of "Raw" on six datasets. These facts demonstrate the efficiency and stability of IV-FS-FRS-2.

VII. CONCLUSION AND FUTURE WORK

To expedite feature selection from large datasets, we study the incremental approach to FS-FRS assuming a dataset can be divided into some sample subsets to be presented sequentially. By an insight into the incremental change of feature subset as sample subsets are presented and processed sequentially, we propose strategies for adding and deleting features based on the updated relative discernibility relations. To exploit the strategies, we design two incremental versions for FS-FRS. One updates the relative discernibility relations and the feature subset as sample subsets arrive sequentially, and outputs the feature subset where there is no sample subset. The other only updates the relative discernibility relations as sample subsets arrive continuously, and then computes the feature subset where no sample subset is left, which is just the feature subset from the whole dataset. The experimental results demonstrate the following facts.

1) Our incremental approach can expedite feature selection based fuzzy rough set.
2) It is possible to employ our incremental approach to handle large datasets.
3) Our second version is more efficient than our first one.

Based on the above results, some further investigations are as follows.

1) The global optimization techniques can find a global optimal feature subset that may improve the accuracy of a classifier to some extent. As suggested, hence, we will focus on the incremental mechanisms for the global optimal method to improve the time efficiency.
2) To overcome the excess storage of fuzzy relation matrices, we will design a novel incremental method for feature selection so that it can efficiently handle big datasets.
3) Considering the feasibility of the true sliding window, we will introduce it into the incremental process of feature subset, in order to further accelerate FS-FRS and improve the accuracy of feature subset.
4) We will investigate how to select the size of the sample subset sequence.

REFERENCES

[1] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: An accelerator for attribute reduction in rough set theory," *Artif. Intell.*, vol. 174, no. 9. pp. 597–618, 2010.
[2] M. Marinaki *et al.*, "A bumble bees mating optimization algorithm for the feature selection problem," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 4, pp. 519–538, 2016.
[3] N. N. Morsi and M. Yakout, "Axiomatics for fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 100, no. 1, pp. 327–342, 1998.
[4] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, pp. 191–209, 1990.
[5] E. C. C. Tsang, D. G. Chen, D. S. Yeung, X. Z. Wang, and J. W. T. Lee, "Attributes reduction using fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1130–1141, Oct. 2008.
[6] D. G. Chen and Y. Y. Yang, "Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set model," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1325–1334, Oct. 2014.
[7] J. Zhang, S. F. Ding, and N. Zhang, "Incremental extreme learning machine based on deep feature embedded," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 1, pp. 111–120, 2016.
[8] A. Ekbal and S. Saha, "Simultaneous feature and parameter selection using multiobjective optimization: Application to named entity recognition," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 4, pp. 597–611, 2016.
[9] R. Jensen and Q. Shen, "Fuzzy-rough attributes reduction with application to web categorization," *Fuzzy Sets Syst.*, vol. 141, no. 3, pp. 469–485, 2004.
[10] R. B. Bhatt and M. Gopal, "On fuzzy rough sets approach to feature selection," *Pattern Recognit.*, vol. 26, no. 7, pp. 965–975, 2005.
[11] Q. H. Hu, D. R. Yu, and Z. X. Xie, "Information-preserving hybrid data reduction based on fuzzy rough techniques," *Pattern Recognit. Lett.*, vol. 27, no. 5, pp. 414–423, 2006.
[12] 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.
[13] D. G. Chen, L. Zhang, S. Y. Zhao, Q. H. Hu, and P. F. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 385–389, Apr. 2012.
[14] C. Z. Wang *et al.*, "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, to be published.
[15] Y. N. Fan *et al.*,"Rule induction based on an incremental rough set," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11439–11450, 2009.
[16] N. Shan and W. Ziarko, "Data-based acquisition and incremental modification of classification rules," *Comput. Intell.*, vol. 11, no. 2, pp. 357–370, 1995.
[17] C. C. Huang, T. L. B. Tseng, Y. N. Fan, and C. H. Hsu, "Alternative rule induction methods based on incremental object using rough set theory," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 372–389, 2013.
[18] Z. Zheng and G. Wang, "RRIA: A rough set and rule tree based incremental knowledge acquisition algorithm," *Fundam. Inf.*, vol. 59, no. 2/3, pp. 299–313, 2004.

[19] J. Blaszczynski and R. Slowinski, "Incremental induction of decision rules from dominance-based rough approximations," *Electron. Notes Theor. Comput. Sci.*, vol. 82, no. 4, pp. 40–51, 2003.
[20] B. Tripathy *et al.*,"Enhanced rule induction using incremental approach for a dynamic information system," *Int. J. Comput. Theory Eng.*, vol. 3, no. 4, pp. 509, 2011.
[21] S. Tsumoto, "A new framework for incremental rule induction based on rough sets," in *Proc. IEEE Int. Conf. Granular Comput.*, 2011, pp. 681–686.
[22] H. Chen, T. R. Li, C. Luo, S. -J. Horng, and G. Wang, "A rough set-based method for updating decision rules on attribute values coarsening and refining," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2886–2899, Dec. 2014.
[23] C. C. Chan, "A rough set approach to attribute generalization in data mining," *Inf. Sci.*, vol. 107, no. 1, pp. 169–176, 1998.
[24] T. Li, D. Ruan, W. Geert, J. Song, and Y. Xu, "A rough sets based characteristic relation approach for dynamic attribute generalization in data mining," *Knowl-Based Syst.*, vol. 20, no. 5, pp. 485–494, 2007.
[25] Y. Cheng, "The incremental method for fast computing the rough fuzzy approximations," *Data Knowl. Eng.*, vol. 70, no. 1, pp. 84–100, 2011.
[26] J. Zhang, T. Li, D. Ruan, and D. Liu, "Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems," *Int. J. Approx. Reason.*, vol. 53, no. 4, pp. 620–635, 2012.
[27] S. Li, T. Li, and D. Liu, "Incrementalupdating approximations in dominance-based rough sets approach under the variation of the attribute set," *Knowl.-Based Syst.*, vol. 40, pp. 17–26, 2013.
[28] C. Luo, T. Li, and H. Chen, "Dynamic maintenance of approximations in set-valued ordered decision systems under the attribute generalization," *Inf. Sci.*, vol. 257, pp. 210–228, 2014.
[29] X. Yang, Y. Qi, H. Yu, X. Song, and J. Yang, "Updating multigranulation rough approximations with increasing of granular structures," *Knowl.-Based Syst.*, vol. 64, pp. 59–69, 2014.
[30] D. Liu, T. R. Li, and J. B. Zhang, "Incremental updating approximations in probabilistic rough sets under the variation of attributes," *Knowl.-Based Syst.*, vol. 73, pp. 81–96, 2015.
[31] H. Chen, T. Li, S. Qiao, and D. Ruan, "A rough set based dynamic maintenance approach for approximations in coarsening and refining attribute values," *Int. J. Intell. Syst.*, vol. 25, no. 10, pp. 1005–1026, 2010.
[32] H. Chen, T. Li, and D. Ruan, "Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining," *Knowl.-Based Syst.*, vol. 31, pp. 140–161, 2012.
[33] C. Luo, T. Li, H. Chen, and L. Lu, "Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values," *Inf. Sci.*, vol. 299, pp. 221–242, 2015.
[34] H. Chen, T. Li, D. Ruan, J. Lin, and C. Hu, "A rough-set-based incremental approach for updating approximations under dynamic maintenance environments," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 2, pp. 274–284, Feb. 2013.
[35] J. B. Zhang *et al.*, "Neighborhood rough sets for dynamic data mining," *Int. J. Intell. Syst.*, vol. 27, no. 4, pp. 317–342, 2012.
[36] S. Li, T. Li, and D. Liu, "Dynamic maintenance of approximations in dominance-based rough set approach under the variation of the object set," *Int. J. Intell. Syst.*, vol. 28, no. 8, pp. 729–751, 2013.
[37] C. Luo, T. Li, H. Chen, and D. Liu, "Incremental approaches for updating approximations in set-valued ordered information systems," *Knowl.-Based Syst.*, vol. 50, pp. 218–233, 2013.
[38] J. Zhang, T. Li, and H. Chen, "Composite rough sets for dynamic data mining," *Inf. Sci.*, vol. 257, pp. 81–100, 2014.
[39] A. P. Zeng, T. R. Li, and L. Chuan, "Incremental approach for updating approximations of gaussian kernelized fuzzy rough sets under variation of object set," *Comp. Sci.*, vol. 40, no. 7, pp. 173–177, 2013.
[40] C. Luo *et al.*, "Matrix approach to decision-theoretic rough sets for evolving data," *Knowl.-Based Syst.*, vol. 99, pp. 123–134, 2016.
[41] F. Wang, J. Liang, and Y. Qian, "Attribute reduction: A dimension incremental strategy," *Knowl.-Based Syst.*, vol. 39, pp. 95–108, 2013.
[42] W. Shu and W. Shu, "Updating attribute reduction in incomplete decision systems with the variation of attribute set," *Int. J. Approx. Reason.*, vol. 55, no. 3, pp. 867–884, 2014.
[43] W. Qian, W. Shu, and C. Zhang, "Feature selection from the perspective of knowledge granulation in dynamic set-valued information system," *J. Inf. Sci. Eng.*, vol. 32, no. 3, pp 783–798, 2016.
[44] A. Zeng, T. Li, D. Liu, J. Zhang, and H. Chen, "A fuzzy rough set approach for incremental feature selection on hybrid information systems," *Fuzzy Sets Syst.*, vol. 258, pp. 39–60, 2015.
[45] F. Wang, J. Wang, and C. Dang, "Attribute reduction for dynamic data sets," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 676–689, 2013.
[46] W. Shu and H. Shen, "Incremental feature selection based on rough set in dynamic incomplete data," *Pattern Recognit.*, vol. 47, no. 12, pp. 3890–3906, 2014.
[47] Z. T. Liu, "An incremental attribute arithmetic for the smallest reduction of attributes," *Acta Electronoca Sin.*, vol. 27, no. 11, pp. 96–98, 1999.
[48] F. Hu, G. Wang, H. Huang, and Y. Wu, "Incremental attribute reduction based on elementary sets," in *Proc. Rough Sets Fuzzy Sets, Data Mining, and Granular-Soft Comput.*, 2005, pp. 185–193.
[49] F. Hu, J. Dai, and G. Y. Wang, "Incremental algorithms for attribute reduction in decision table," *Control Decis.*, vol. 22, no. 3, pp. 268, 2012.
[50] M. Yang, "An incremental updating algorithm for attribute reduction based on improved discernibility matrix," *Chin. J. Comput.*, vol. 30, no. 5, pp. 815–822, 2007.
[51] D. Chen, Y. Yang, and Z. Dong, "An incremental algorithm for attribute reduction with variable precision rough sets," *Appl. Soft Comput.*, vol. 45, pp. 129–149, 2016.
[52] 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.
[53] Y. Yang, D. Chen, and H. Wang, "Active sample selection based incremental algorithm for attribute reduction with rough sets," *IEEE Trans. Fuzzy Syst.*, to be published.
[54] 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.
[55] D. G. Chen and S. Y. Zhao, "Local reduction of decision system with fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 161, no. 13, pp. 1871–1883, 2010.
[56] S. Zhao and E. C. Tsang, "On fuzzy approximation operators in attribute reduction with fuzzy rough sets," *Inf. Sci.*, vol. 178, no. 16, pp. 3163–3176, 2008.
[57] D. Chen, S. Zhao, L. Zhang, and X. Zhang, "Sample pair selection for attribute reduction with rough set," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 11, pp. 2080–2093, Nov. 2012.
[58] D. S. Yeung, D. Chen, E. C. C. Tsang, J. W. T. Lee, and W. Xizhao, "On the generalization of fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343–361, Jun. 2005.
[59] J. Mi and W. Zhang, "An axiomatics characterization of a fuzzy generalization of rough sets," *Inf. Sci.*, vol, 160, pp. 235–249, 2004.
[60] N. Li and J. Y. Xie, "A feature subset selection algorithm based on neighborhood rough set for incremental updating datasets," *Comput. Technol. Dev.*, vol. 21, no. 11, pp. 149–155, 2011.
[61] X. Zhang, C. L. Mei, D. G. Chen, and J. H. Li, "Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy," *Pattern Recognit.*, vol. 56, pp. 1–15, 2016.
[62] R. R. Chhikara and P. Sharma, "A hybrid feature selection approach based on improved PSO and filter approaches for image steganalysis," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 6, pp. 1195–1206, 2016.
[63] E. C. C. Tsang *et al.*, "Feature and instance reduction for PNN classifiers based on fuzzy rough sets," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 1, pp. 1–11, 2016.

Yanyan Yang received the B.Sc. degree in applied mathematics from North University of China, Taiyuan, China, in 2010, and the M.Sc. degree in applied mathematics from North China Electric Power University, Beijing, China, in 2013, where she is currently working toward the Ph.D. degree in engineering.

In October 2015, she spent one year visiting the University of Ulster, Coleraine, U.K., as a Joint Ph.D. student sponsored by The China Scholarship Council. Her current research interests include rough sets, fuzzy sets, and machine learning.

Degang Chen received the M.Sc. degree in science from the Northeast North University, Changchun, China, in 1994, and the Ph.D. degree in science from the Harbin Institute of Technology, Harbin, China, in 2000.

He was a Postdoctoral Fellow with Xían Jiaotong University, Xían, China, from 2000 to 2002, and with Tsinghua University, Beijing, China, from 2002 to 2004. Since 2006, he has been a Professor with the North China Electric Power University, Beijing, China. He has authored or coauthored more than 140 research publications. His research interests include fuzzy groups, fuzzy algebra, fuzzy analysis, rough sets, machine learning.

Hui Wang received the B.Sc. degree in computer science and the M.Sc. degree in artificial intelligence from Jilin University of China, Changchun, China, in 1985 and 1988, respectively, and the D.Phil. degree in informatics from the University of Ulster, Coleraine, U.K., in 1996.

From 1988 to 1992, he was with Jilin University as a Lecturer. He is currently a Professor of computer science and the Head of Artificial Intelligence and Applications Research Group, University of Ulster. His research interests include machine learning, logics and reasoning, combinatorial data analytics, and their applications in image, video, spectra and text analysis. He has more than 200 publications in these areas.

Xizhao Wang (M'03–SM'04–F'12) received the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in September 1998.

From 1998 to 2001, he was a Research Fellow with the Department of Computing, Hong Kong Polytechnic University. From 2001 to 2014, he was with Hebei University as a Professor and the Dean of the School of Mathematics and Computer Sciences. He was the Founding Director of the Key Laboratory on Machine Learning and Computational Intelligence in Hebei Province. Since 2014, he has been a Professor in the Big Data Institute, Shenzhen University, Shenzhen, China. He has edited more than 10 special issues and published 3 monographs, 2 textbooks, and more than 200 peer-reviewed research papers. By the Google scholar, the total number of citations is more than 5000 and the maximum number of citation for a single paper is more than 200. He is on the list of Elsevier 2015/2016 Most Cited Chinese Authors. As a Principle Investigator (PI) or Co-PI, he has completed more than 30 research projects. He has supervised more than 100 M.Phil. and Ph.D. students. His major research interests include uncertainty modeling and machine learning for big data.

Dr. Wang is the Chair of the IEEE Systems, Man, and Cybernetics (SMC) Technical Committee on Computational Intelligence, an Editor-in-Chief of the *Machine Learning and Cybernetics Journal*, and an Associate Editor for a couple of journals in related areas. He was a BoG member of the IEEE SMC Society. He received the IEEE SMCS Outstanding Contribution Award in 2004 and the IEEE SMCS Best Associate Editor Award in 2006. He is the General Co-Chair of the 2002–2017 International Conferences on Machine Learning and Cybernetics, cosponsored by IEEE SMCS. He was a Distinguished Lecturer of the IEEE SMCS.