Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 258 (2015) 39–60

FUZZY sets and systems

www.elsevier.com/locate/fss

A fuzzy rough set approach for incremental feature selection on hybrid information systems

Anping Zeng [a,b], Tianrui Li [a,*], Dun Liu [c], Junbo Zhang [a], Hongmei Chen [a]

[a] *School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China*
[b] *School of Computer and Information Engineering, Yibin University, Yibin 644007, China*
[c] *School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China*

Received 3 September 2013; received in revised form 7 August 2014; accepted 22 August 2014

Available online 6 September 2014

Abstract

In real-applications, there may exist many kinds of data (*e.g.*, boolean, categorical, real-valued and set-valued data) and missing data in an information system which is called as a Hybrid Information System (HIS). A new Hybrid Distance (HD) in HIS is developed based on the value difference metric, and a novel fuzzy rough set is constructed by combining the HD distance and the Gaussian kernel. Considering the information systems often vary with time, the updating mechanisms for attribute reduction (feature selection) are analyzed with the variation of the attribute set. Fuzzy rough set approaches for incremental feature selection on HIS are presented. Then two corresponding incremental algorithms are proposed, respectively. Finally, extensive experiments on eight datasets from UCI and an artificial dataset show that the incremental approaches significantly outperform non-incremental approaches with feature selection in the computational time.

Keywords: Fuzzy rough sets; Incremental learning; Feature selection; Hybrid information systems; Big data

1. Introduction

Feature selection is an important technique in pattern recognition, machine leaning and data mining as there usually are many candidate attributes collected to represent recognition problems. In some real-world applications, data expand quickly, and tens, hundreds even thousands of attributes are stored in databases. These attributes may be of different types (*e.g.*, boolean, categorical, real-valued and set-valued data). The information system that includes different types of attributes is called as a Hybrid Information System (HIS). In HIS, some of attributes are irrelevant to the learning or recognition tasks. Experiments have shown irrelative attributes will deteriorate the performance of the learning algorithms for the cause of dimensionality, which increase training and testing times [1,2]. Feature selection

* Corresponding author.
E-mail addresses: zengap@126.com (A. Zeng), trli@swjtu.edu.cn (T. Li), newton83@163.com (D. Liu), junbozhang86@163.com (J. Zhang), hmchen@swjtu.edu.cn (H. Chen).

http://dx.doi.org/10.1016/j.fss.2014.08.014
0165-0114/

is an effective approach to remove the irrelative attributes for big data analysis. In recent years, feature selection in the HIS has received much attention. Many optimization algorithms of feature selection were presented to increase the classification accuracy in HIS [3–8].

The concept of feature selection (*aka.* attribute reduction), is viewed as one of the most important techniques in Rough Set Theory (RST) [9]. However, it is hard to use the traditional RST in handling hybrid attributes (*e.g.*, real-valued attributes). Discretization of the real-valued attributes is one way to solve this problem, but it may cause information loss. Another approach is the use of Fuzzy Rough Sets (FRS). FRS encapsulate the related but distinct concepts of fuzziness and indiscernibility, both occur as a result of uncertainty existing in knowledge [10–15].

Feature selection via FRS was first proposed in [16,17] which considers the problem of evaluating the hypoxic resistance of a patient on the basis of the values of his blood pressure during a barocamera examination. The measurements were evaluated by the FRS criterion. In [18], fuzzy-rough attribute reduction was proposed. The dependency function to measure the importance of attributes by FRS was defined and an algorithm to compute a reduct was designed. This algorithm was tested with some practical data sets from web categorization and was claimed to have better performance. In [19], Chen et al. introduced the concept of local reduction with FRS for decision systems. The local reduction can identify key conditional attributes and offer a minimal description for every single decision class. In [20], the FRS was used to compute both the relevance and significance of features. In [21,22], Hu et al. proposed Gaussian kernel based FRS and applied it into feature selection. In this model, Gaussian kernels were first introduced to acquire fuzzy relations between samples described by numeric attributes.

Nowadays we live in a world of big data [23–25]. Big data has the characteristics of 4Vs, *i.e.*, Volume, Variety, Velocity and Value [26]. Volume means the amount of data that needs to be handled is very large, and exabytes, zettabytes, and even higher amounts of data are described in big data applications. Variety means that the data is varied in nature, and structured, unstructured or semi-structured data needs to be properly combined to make the most of the analysis. Velocity means that a high rate of sampling is common in big data problems. Value means high yield will be achieved if the big data is used reasonable by analyzing correctly and accurately. Considering the velocity characteristic, algorithms which can improve efficiency of knowledge discovery are badly needed. Incremental learning is essential for the knowledge discovery which is an important manner of human intelligence. Especially in two cases, the first one is when the dataset cannot be collected at one time and the batch computing cannot be carried out, *e.g.* online applications [23] or interaction query [24]; the second case is when the data set is too big and it cannot be calculated at one turn due to the limitation of the computation capability and the memory size. The data set has to be cut to bulk and added in succession. Incremental learning is feasible when the data structure and information of the previous data is stored and the old data need not to be scanned again. Then the relationships between the new data added and the data structure stored are analyzed in order to require new knowledge. Therefore, in real-life applications, information systems may be very large [27–30] and vary with time. In FRS, the generation of fuzzy relations among samples is often very costly or even intractable. In fact, not only in FRS but also in other RST extensions, more and more computing problems arise due to the appearance of big data. In our study, the dynamic properties and the big size of the data set are both taken into consideration. Incremental learning method is studied which aims to improve the efficiency of feature selection under fuzzy rough set. Incremental learning which is fully used of the information get previously can improve the efficiency of knowledge discovery in big data.

The variation of objects has been widely considered in incremental learning. Shan et al. presented a discernibility-matrix based incremental methodology to find all maximally generalized rules [31]. Bang et al. proposed an incremental inductive learning algorithm to find a minimal set of rules for a decision table without recalculating all the set of instances when a new instance is added into the universe [32]. Liu et al. proposed an incremental approach for inducing interesting knowledge when the object set varies over time [33]. Zheng et al. presented an effective incremental approach for knowledge acquisition based on the rule tree [34]. Wang et al. constructed an incremental rule acquisition algorithm based on Variable Precision Rough Sets (VPRS) while inserting new objects into the information system [35]. Zhang et al. proposed an incremental rule acquisition algorithm based on neighborhood rough sets when the object set evolves over time [36]. Li et al. proposed a dynamic maintenance of approximations in Dominance-based Rough Sets Approach (DRSA) when adding or deleting objects in the universe [37]. In addition, Luo et al. proposed an incremental approaches for updating approximations in set-valued ordered information systems [38]. Zhang et al. proposed a composite rough set model for dynamic data mining [39]. In FRS, Zeng et al. proposed an incremental approach for updating approximations of Gaussian kernelized FRS under the variation of the object set [40]. In attribute reduction, several incremental reduction algorithms have been proposed to deal with dynamic data sets. A common character of these algorithms is that they were only applicable when new data are generated one by one [41–43], whereas many real data from applications are generated in groups. Liang et al. introduced incremental mechanisms for three representative information entropies and then developed a group incremental rough feature selection algorithm based on information entropy [44].

Under the variation of the attribute set, Chan proposed an incremental method for updating the approximations of a concept based on the lower and upper boundary sets [45]. Li et al. proposed a generalized rough set approach to attribute generalization in data mining [46], and then a rough sets based characteristic relation approach for dynamic attribute generalization was proposed [47]. In Rough Fuzzy Sets, Cheng proposed approaches for incremental updating approximations when the attribute set evolves over time [48]. In addition, Zhang et al. discussed the strategies and propositions under VPRS when attributes are changed [49]. Li et al. introduced a dominance matrix to calculate P-dominating sets and P-dominated sets in DRSA, and proposed the incremental algorithms for updating approximations of DRSA when the attribute set varies [50]. Luo et al. focused on maintaining approximations dynamically in set-valued ordered decision systems under the attribute generalization [51]. However, there is no report about the incremental learning based on FRS under the variation of the attribute set so far. In this paper, we discuss the principles of incremental feature selection when the attributes change (adding or deleting) dynamically under FRS at first. Then we propose two incremental feature selection algorithms. Finally, the performances of two incremental algorithms are evaluated on several UCI datasets.

The rest of this paper is organized as follows. In Section 2, some preliminaries are introduced. In Section 3, the generating methods of fuzzy information granules in HIS are presented. In Section 4, attribute evaluation and feature selection with FRS are discussed. In Section 5, the incremental updating principles for feature selection are analyzed under the variation of the attribute set. In Section 6, some experiments are designed to evaluate the performance. In Section 7, we conclude the paper.

2. Fundamentals on Pawlak's rough sets and FRS

In this section, we briefly introduce basic concepts of rough sets and FRS [9,10,52–54].

2.1. Pawlak's rough sets

Definition 1. An information system is defined as a pair $\langle U, A\rangle$, where U is a non-empty finite set of objects, $A = C \cup D$ is a non-empty finite set of attributes, C denotes the set of condition attributes and D denotes the set of decision attributes, $C \cap D = \emptyset$. Each attribute $a \in A$ is associated with a set V_a of its value, called the domain of a.

The RST describes a crisp subset of a universe by two definable subsets called lower and upper approximations. By using the lower and upper approximations, the knowledge hidden in information systems can be discovered and expressed in the form of decision rules.

Definition 2. Let U be a finite universe, $U \neq \emptyset$. R is an equivalence relation on U if for $\forall x, y, z \in U$, we have:

(1) Reflexivity: $R(x, x) = 1$;
(2) Symmetry: $R(x, y) = R(y, x)$;
(3) Transitivity: $R(x, y) = R(y, z) \Rightarrow R(x, y) = R(x, z)$.

The equivalence relation partitions the universe into a family of disjoint subsets, called equivalence classes. The equivalence class including x is denoted by $[x]_R$.

Definition 3. Let (U, R) be a Pawlak approximation space. The universe $U \neq \emptyset$. $R \subseteq U \times U$ is an equivalence relation on U. U/R denotes the family of all equivalence classes R, and $[x]_R$ denotes an equivalence class of R containing an element $x \in U$. For any $X \subseteq U$, the lower and the upper approximations of X are defined respectively as follows:

$$\begin{cases} \underline{R}X = \{x \in U \mid [x]_R \subseteq X\}; \\ \overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \end{cases} \tag{1}$$

2.2. Fuzzy rough sets

Definition 4. (See [53].) Let U be a non-empty and finite universe, and $\tilde{A}$ be a fuzzy subset on U. $\forall u \in U$, $\tilde{A}$ is defined by the following real function:

$$\begin{aligned} \mu_{\tilde{A}}: \quad & U \to [0,1] \\ & u \mapsto \mu_{\tilde{A}}(u), \end{aligned} \tag{2}$$

where $\mu_{\tilde{A}}(u)$ (shortly, $\tilde{A}(u)$) is the membership of $\tilde{A}$.

Let $U = \{x_1, x_2, ..., x_n\}$. A fuzzy set $\tilde{A}$ can be represented as $\{\tilde{A}(x_1)/x_1, \tilde{A}(x_2)/x_2, ..., \tilde{A}(x_n)/x_n\}$.

Definition 5. Let U be a finite universe, $U \neq \emptyset$, and $R \in F(U \times U)$. R is a fuzzy equivalence relation on U if for $\forall x, y, z \in U$, we have:

(1) Reflexivity: $R(x,x) = 1$;
(2) Symmetry: $R(x,y) = R(y,x)$;
(3) Min–max transitivity: $min(R(x,y), R(y,z)) \leq R(x,z)$.

A binary operator T on the unit interval $I = [0,1]$ is said to be a triangular norm. $\forall a, b, c \in I$, the following conditions are satisfied:

(1) Commutativity: $T(a,b) = T(b,a)$;
(2) Associativity: $T(T(a,b),c) = T(a, T(b,c))$;
(3) Monotonicity: $a \leq c, b \leq d \Rightarrow T(a,b) \leq T(c,d)$, and
(4) Boundary condition: $T(a,1) = a, T(1,a) = a$.

We say that a binary operator S on the unit interval I is a triangular-conorm (shortly, t-conorm or s-norm) if it is increasing, associative, and commutative and satisfies the boundary condition of the form $S(a,0) = a, \forall a \in [0,1]$.

A negator N is a decreasing mapping from $[0,1] \to [0,1]$, satisfying $N(0) = 1$ and $N(1) = 0$. The negator $N_s(a) = 1 - a$ is usually referred to as the standard negator. A negator N is said to be involutive if $N(N(a)) = a, \forall a \in [0,1]$. If $S(a,b) = N(T(N(a), N(b))), T(a,b) = N(S(N(a), N(b))), \forall a, b \in [0,1]$, we say T and S are dual with respect to N.

It is clear that $T_{cos}(a,b) = max\{ab - \sqrt{1-a^2}\sqrt{1-b^2}, 0\}$ is a T-norm, and its dual $S_{cos}(a,b) = min\{a + b - ab + \sqrt{2a - a^2}\sqrt{2b - b^2}, 1\}$.

Definition 6. Let U be a finite universe, $U \neq \emptyset$, and $R \in F(U \times U)$. R is a T-fuzzy equivalence relation on U if $\forall x, y, z \in U$, we have:

(1) Reflexivity: $R(x,x) = 1$;
(2) Symmetry: $R(x,y) = R(y,x)$;
(3) T-transitivity: $T(R(x,y), R(y,z)) \leq R(x,z)$.

Since T_{cos}-norm is a T-norm, we have $T_{cos}(R(x,y), R(y,z)) \leq R(x,z)$. Therefore, R is a T_{cos}-fuzzy equivalence relation on U.

Definition 7. Let R be a fuzzy equivalence relation on U and X be a fuzzy subset of U. The fuzzy lower and upper approximations of X were defined respectively as follows.

$$\begin{cases} \underline{R}X(x) = \inf_{y \in U} \{max(1 - R(x,y), X(y))\}; \\ \overline{R}X(x) = \sup_{y \in U} \{min(R(x,y), X(y))\}. \end{cases} \tag{3}$$

Table 1
A hybrid information system.

U	Headache(a_1)	Muscle pain(a_2)	Temperature(a_3)	Syndrome(a_4)	d
x_1	Sick	Yes	40	{C, R, A}	Flu
x_2	Sick	Yes	39.5	{C, R, A}	Flu
x_3	Middle	?	39	{C}	Flu
x_4	Middle	Yes	36.8	{R}	Rhinitis
x_5	Middle	No	?	{R}	Rhinitis
x_6	No	No	36.6	{R, A}	Health
x_7	No	?	?	{A}	Health
x_8	No	Yes	38	{C, R, A}	Flu
x_9	?	Yes	37	{R}	Health

More generally, Yeung et al. proposed a model of FRS with a pair of t-norm T and S-norm in [55].

$$\begin{cases} \underline{R}X(x) = \inf_{y\in U}\big\{S\big(N\big(R(x,y)\big), X(y)\big)\big\}; \\ \overline{R}X(x) = \sup_{y\in U}\big\{T\big(R(x,y), X(y)\big)\big\}. \end{cases} \tag{4}$$

3. Generation of fuzzy information granules in HIS

Definition 8. A HIS can be written as $\langle U, C \cup D, V, f\rangle$, where U is the set of objects, $C = C^r \cup C^s \cup C^c \cup C^b$, C^r is the real-valued attribute set, C^s is the set-valued attribute set, C^c is the categorical attribute set and C^b is the boolean attribute set. D denotes the set of decision attributes. $C^r \cap C^s = \emptyset$, $C^r \cap C^c = \emptyset$, $C^s \cap C^c = \emptyset$, $C^r \cap C^b = \emptyset$, $C^s \cap C^b = \emptyset$, $C^c \cap C^b = \emptyset$ and $C \cap D = \emptyset$.

Example 1. Table 1 is a HIS with a categorical attribute “Headache”, a boolean attribute “Muscle pain”, a real-valued attribute “Temperature”, a set-valued attribute “Syndrome” (denoted as a_1, a_2, a_3, a_4, respectively), and a decision attribute d. “?” denotes the unknown value.

3.1. Hybrid distance

In HIS, there may be different types of attributes (*e.g.*, boolean, categorical, real-valued, and set-valued). To construct the distance among objects efficiently, a novel distance function should be presented.

In order to deal with the value difference under the boolean attributes, the value difference is defined as follows.

Definition 9. Let $HIS = \langle U, C \cup D, V, f\rangle$. $\forall x, y \in U$, $\forall a \in C$ and a is a boolean attribute. $a(x)$, $a(y)$ denote the values of x, y under a, respectively. The value difference of boolean (vdb) attributes is as follows.

$$vdb\big(a(x), a(y)\big) = \begin{cases} 0, & a(x) = a(y) \\ 1, & a(x) \neq a(y) \end{cases} \tag{5}$$

In order to deal with the value difference under the categorical attributes, Wilson et al. [58] introduced a value difference metric (vdm). Based on it, the normalized value difference under the categorical attributes is defined as follows:

Definition 10. Let $HIS = \langle U, C \cup D, V, f\rangle$. $\forall x, y \in U$, $\forall a \in C$ and a is a categorical attribute. $a(x)$, $a(y)$ denote the values of x, y under a, respectively. The value difference of categorical attributes is defined as follows.

$$vdm\big(a(x), a(y)\big) = \sqrt{\frac{1}{|U/D|}\sum_{d_i \in U/D}\left(\frac{|a(x)\cap d_i|}{|a(x)|} - \frac{|a(y)\cap d_i|}{|a(y)|}\right)^2}, \tag{6}$$

where $|.|$ denotes the support number, $|U/D|$ denotes the number of decision and it is clear that $vdm(a(x), a(y)) \in [0, 1]$.

In [58], Wilson et al. defined a value difference under real-valued attributes.

Definition 11. Let $HIS = \langle U, C \cup D, V, f\rangle$. $\forall x, y \in U, \forall a \in C$ and a is a real-valued attribute. $a(x)$, $a(y)$ denote the values of x, y under a, respectively. The value difference of real-valued (vdr) attributes is defined as follows.

$$vdr\big(a(x), a(y)\big) = \frac{|a(x) - a(y)|}{4\delta_a} \tag{7}$$

where δ_a is the standard deviation under the attribute a.

In order to deal with the unknown values (denoted by "?"), Wilson et al. also defined a value difference as follows [58].

Definition 12. Let $HIS = \langle U, C \cup D, V, f\rangle$. $\forall x, y \in U, \forall a \in C$. $a(x)$, $a(y)$ denote the values of x, y under a, respectively. If $a(x) = ?$ or $a(y) = ?$ and $x \neq y$, the value difference of them is defined as follows.

$$vdu\big(a(x), a(y)\big) = 1. \tag{8}$$

According to Definition 12, the value difference can be set as 1 between an unknown value and another one.

A set-valued attribute can be treated as a set of boolean attributes. For example, attribute a_4 in Table 1, the subset which has maximum cardinal number in the domain V_d is {C, R, A}. Therefore, attribute a_4 can be divided into three boolean attributes (C, R, A, respectively). Therefore, set-valued $\{C, R, A\} = \{C = Yes,\ R = Yes,\ A = Yes\}$, $\{C, R\} = \{C = Yes,\ R = Yes,\ A = ?\}$. Because the value difference between "?" and other values is equal to 1, the value difference between {C, R, A} and {C, R} is $1/3$. The value difference of set-valued attributes is defined as follows:

Definition 13. Let $HIS = \langle U, C \cup D, V, f\rangle$. $\forall x, y \in U, \forall a \in C$ and a is a set-valued attribute. Let V_a be the domain of a. $a(x)$, $a(y)$ denote the values of x, y under a, respectively. The value difference of set-valued (vds) attributes is defined as follows.

$$vds\big(a(x), a(y)\big) = 1 - \frac{|a(x) \cap a(y)|}{s} \tag{9}$$

where s is the maximum cardinal number (cardinality) in the subset of V_a.

In order to deal with the hybrid and incomplete attributes, according to the above five Definitions, a novel Hybrid Distance (HD) is defined as follows:

Definition 14. Given a HIS, the HD is defined as:

$$HD(x, y) = \sqrt{\sum_{a=1}^{m} d^2\big(a(x), a(y)\big)} \tag{10}$$

where m is the number of attributes, and

$$d\big(a(x), a(y)\big) = \begin{cases} 1, & a(x) = ? \text{ or } a(y) = ? \text{ and } x \neq y \\ vdb(a(x), a(y)), & a \text{ is } a \text{ boolean attribute} \\ vdm(a(x), a(y)), & a \text{ is } a \text{ categorical attribute} \\ vds(a(x), a(y)), & a \text{ is } a \text{ set-valued attribute} \\ vdr(a(x), a(y)), & a \text{ is } a \text{ real-valued attribute} \end{cases} \tag{11}$$

Example 2. Based on Example 1, we can compute the HD matrix. According to formula (10), the following results hold:

(1) Because attribute a_1 is categorical, $d(a_1(x_1), a_1(x_3)) = vdm(a_1(x_1), a_1(x_3)) = \sqrt{\frac{1}{3}((\frac{2}{2} - \frac{1}{3})^2 + (\frac{0}{2} - \frac{2}{3})^2 + (\frac{0}{2} - \frac{0}{3})^2)} = 0.54$.

(2) $a_2(x_3) = ?$, therefore, $d(a_2(x_1), a_2(x_3)) = 1$.
(3) Since attribute a_3 is real-valued, $d(a_3(x_1), a_3(x_3)) = \frac{a_3(x_1)-a_3(x_3)}{4\delta_{a_3}} = (40-39)/(4 \times 1.28) = 0.19$.
(4) Because attribute a_4 is set-valued, $d(a_4(x_1), a_4(x_3)) = vds(a_4(x_1), a_4(x_3)) = 1 - \frac{1}{3} = 0.67$.

Therefore,

$$HD(x_1, x_3) = \left(\sum_{a=1}^{4} d^2\big(a(x), a(y)\big)\right)^{1/2} = \sqrt{0.54^2 + 1^2 + 0.19^2 + 0.67^2} = 1.33.$$

3.2. Generation of fuzzy information granules with hybrid attributes

In [22], Hu et al. combined Gaussian kernel function with Euclidean distance and generated fuzzy information granules. The Euclidean distance does not deal with hybrid attributes directly. Therefore, we replace Euclidean distance with the HD distance and construct fuzzy information granules.

Let U be a finite universe and $U \neq \emptyset$. The samples are m-dimension vectors. $\forall x_i, x_k \in U$, $x_i = \langle x_{i1}, x_{i2}, ..., x_{im}\rangle$, $x_k = \langle x_{k1}, x_{k2}, ..., x_{km}\rangle$. The Gaussian kernel function

$$k(x_i, x_k) = \exp\left(-\frac{\|x_i - x_k\|^2}{2\delta^2}\right) \tag{12}$$

can be used to compute the similarity between samples x_i and x_k. $\|x_i - x_k\|$ is the HD distance between x_i and x_k. We have

(1) $k(x_i, x_k) \in [0, 1]$;
(2) $k(x_i, x_k) = k(x_k, x_i)$;
(3) $k(x_i, x_i) = 1$.

Theorem 1. *(See [56].) Gaussian kernel $k : U \times U \to [0, 1]$ satisfies T_{cos}-transitive, and the fuzzy relation R_G computed with Gaussian kernel is a T_{cos}-equivalence relation.*

Example 3. Based on Table 1, we use Gaussian kernel function to compute the T_{cos}-equivalence relation R_G. Let $\delta^2 = 0.8$. Each sample is a 4-D vector. The fuzzy relation between each two samples can be computed by formula (12). For example, $R_G(x_1, x_3) = k(x_1, x_3) = \exp(-\frac{HD^2(x_1,x_3)}{2\times 0.8}) = \exp(-1.53^2/1.6) = 0.23$. Therefore,

$$R_G = \begin{pmatrix} 1 & 0.99 & 0.23 & 0.37 & 0.17 & 0.33 & 0.13 & 0.54 & 0.36 \\ & 1 & 0.23 & 0.39 & 0.17 & 0.34 & 0.13 & 0.55 & 0.37 \\ & & 1 & 0.38 & 0.22 & 0.27 & 0.13 & 0.23 & 0.20 \\ & & & 1 & 0.38 & 0.22 & 0.13 & 0.42 & 0.54 \\ & & & & 1 & 0.17 & 0.13 & 0.17 & 0.21 \\ & & & & & 1 & 0.27 & 0.64 & 0.21 \\ & & & & & & 1 & 0.22 & 0.12 \\ & & & & & & & 1 & 0.40 \\ & & & & & & & & 1 \end{pmatrix}.$$

Given a hybrid attribute set $B \subseteq C$, the algorithm for generating fuzzy information granules is designed as follows (see Algorithm 1).

The time complexity of Algorithm 1 is $O(m \times n^2)$, where m and n are the numbers of features and samples, respectively.

3.3. Gaussian kernelized FRS

Since T_{cos} is a special T-norm, according to Theorem 1, the following definition can be obtained.

Algorithm 1: Fuzzy Relation Computing based on Gaussian Kernel (FRC-GK).

```
Input: The universe U, and the attribute set B.
Output: The fuzzy relation R_G under B.
for i = 0; i < |U|; i++ do
    for j = |U| - 1; j > i; j-- do
        vda = 0;
        for each a ∈ B do
            vda += d^2(a(x_i), a(x_j));          // Computing value differences.
        end
        Compute HD(x_i, x_j) = sqrt(vda);
        R_G(x_i, x_j) = exp(-HD^2(x_i, x_j) / (2 × δ^2));
    end
end
```

Definition 15. Let R_G be a Gaussian Kernelized T_{cos}-fuzzy equivalence relation on U and X be a fuzzy subset of U. The fuzzy lower and upper approximations of X are defined respectively as follows.

$$\begin{cases} \underline{R_G}X(x) = \inf_{y\in U} S_{cos}\big(N\big(R_G(x,y)\big), X(y)\big); \\ \overline{R_G}X(x) = \sup_{y\in U} T_{cos}\big(R_G(x,y), X(y)\big) \end{cases} \tag{13}$$

where $R_G(x, y) = k(x, y)$.

Let $HIS = \langle U, C \cup D, V, f\rangle$. $R/D = \{d_i\}, i = 1, 2, ..., |R/D|$. We have the following relationship $\forall x \in d_i$, $d_i(x) = 1$; otherwise, $d_i(x) = 0$. Therefore, we can approximate the decision regions with the fuzzy granules induced by Gaussian function.

Proposition 1. *(See [21].) $HIS = \langle U, C \cup D, V, f\rangle$, $\forall d_i \in R/D$,*

$$\begin{cases} \underline{R_G}d_i(x) = \inf_{y\notin d_i} \sqrt{1 - {R_G}^2(x,y)}; \\ \overline{R_G}d_i(x) = \sup_{y\in d_i} R_G(x,y). \end{cases} \tag{14}$$

To simple the computing, we can generate the fuzzy lower and upper approximations as follows:

Proposition 2. *$HIS = \langle U, C \cup D, V, f\rangle$, $\forall x \in U, \forall d_i \in R/D$.*

$$\begin{cases} \underline{R_G}d_i(x) = \sqrt{1 - \Big(\sup_{y\notin d_i} R_G(x,y)\Big)^2}; \\ \overline{R_G}d_i(x) = \sup_{y\in d_i} R_G(x,y). \end{cases} \tag{15}$$

Note, it is clear that $y = \sqrt{1-x^2}, x \in [0, 1]$ is a monotonically decreasing function. Because $R_G(x, y) \in [0, 1]$, $\sqrt{1 - (\sup(R_G(x, y)))^2} = \inf(\sqrt{1 - {R_G}^2(x, y)})$ holds. Therefore, $\underline{R_G}d_i(x) = \sqrt{1 - (\sup_{y\notin d_i} R_G(x, y))^2}$.

Example 4. Based on Examples 1 and 3, $R/D = \{d_1, d_2, d_3\}, d_1 = \{x_1, x_2, x_3, x_8\},\ d_2 = \{x_4, x_5\}, d_3 = \{x_6, x_7, x_9\}$. According to Proposition 2,

$$\begin{aligned} \underline{R_G}d_1(x_1) &= \sqrt{1 - \Big(\sup_{y\notin d_1} R_G(x_1,y)\Big)^2} \\ &= \sqrt{1 - \big(\sup\big\{R_G(x_1,x_4), R_G(x_1,x_5), R_G(x_1,x_6), R_G(x_1,x_7), R_G(x_1,x_9)\big\}\big)^2} \\ &= \sqrt{1 - \big(\sup\{0.37, 0.17, 0.33, 0.13, 0.36\}\big)^2} \\ &= 0.93. \end{aligned}$$

Similarly, we have,

$$\underline{R_G}d_1 = \{0.93/x_1, 0.92/x_2, 0.92/x_3, 0/x_4, 0/x_5, 0/x_6, 0/x_7, 0.77/x_8, 0/x_9\}.$$
$$\underline{R_G}d_2 = \{0/x_1, 0/x_2, 0/x_3, 0.84/x_4, 0.97/x_5, 0/x_6, 0/x_7, 0/x_8, 0/x_9\}.$$
$$\underline{R_G}d_3 = \{0/x_1, 0/x_2, 0/x_3, 0/x_4, 0/x_5, 0.77/x_6, 0.98/x_7, 0/x_8, 0.84/x_9\}.$$
$$\begin{aligned}\overline{R_G}d_1(x_1) &= \sup_{y\in d_1} R_G(x_1, y) \\ &= \sup\{R_G(x_1,x_1), R_G(x_1,x_2), R_G(x_1,x_3), R_G(x_1,x_8)\} \\ &= \sup\{1, 0.99, 0.23, 0.54\} = 1.\end{aligned}$$

Similarly, the other upper approximations can be computed. Therefore, we have

$$\overline{R_G}d_1 = \{1/x_1, 1/x_2, 1/x_3, 0.42/x_4, 0.22/x_5, 0.64/x_6, 0.22/x_7, 1/x_8, 0.4/x_9\}.$$
$$\overline{R_G}d_2 = \{0.37/x_1, 0.39/x_2, 0.38/x_3, 1/x_4, 1/x_5, 0.22/x_6, 0.13/x_7, 0.42/x_8, 0.54/x_9\}.$$
$$\overline{R_G}d_3 = \{0.36/x_1, 0.37/x_2, 0.27/x_3, 0.54/x_4, 0.21/x_5, 1/x_6, 1/x_7, 0.64/x_8, 1/x_9\}.$$

4. Attribute evaluation and feature selection with FRS

Definition 16. (See [22].) Given a decision table $\langle U, C \cap D\rangle$, R_G is a T-equivalence relation on U computed with Gaussian kernel in the feature space $B \subseteq C$. U is divided into $\{d_1, d_2, ..., d_I\}$ with the decision attribute. The fuzzy positive regions of D in terms of B are defined as follows:

$$POS_B(D) = \bigcup_{i=1}^{I} \underline{R_G}d_i. \tag{16}$$

Positive region of D is a fuzzy set, and the membership of a sample to the positive regions of decision reflects the degree of the sample necessarily belongs to its decision class. The higher the membership is, the more certain the classification outcome is.

Definition 17. (See [22].) Given a $HIS = \langle U, C \cup D\rangle$, R_G is a T_{cos}-equivalence relation on U computed with Gaussian kernel in the feature space $B \subseteq C$. U is divided into $\{d_1, d_2, ..., d_I\}$ with the decision attribute. The fuzzy positive regions of D in terms of B are given by $\bigcup_{i=1}^{I} \underline{R_G}d_i$. The quality of approximating classification is defined as follows:

$$\gamma_B(D) = \frac{|\bigcup_{i=1}^{I} \underline{R_G}d_i|}{|U|}, \quad \text{where } \bigcup_{i=1}^{I} \underline{R_G}d_i = \sum_i \sum_{x\in d_i} \underline{R_G}d_i(x). \tag{17}$$

In [22], Hu et al. proposed a computation algorithm (DGKA) of fuzzy dependency (see Algorithm 2). By considering DGKA does not include the process of computing fuzzy relation, the time complexity is $m \times n \log n$, where m and n are the numbers of features and samples, respectively.

Followed by the Pawlak rough set model [9], we can define the concepts of redundancy, indispensability and reducts of decision systems based on the fuzzy dependency.

Definition 18. (See [22].) Given a $HIS = \langle U, C \cup D\rangle$, $a \in B \subseteq C$. If $\gamma_B(D) = \gamma_{B-a}(D)$, we say a is redundant in B with respect to D; otherwise, we say a is indispensable in B to D.

Definition 19. (See [22].) Given a $HIS = \langle U, C \cup D\rangle$, $B \subseteq C$. We say B is a relative reduct to D if B satisfies the following conditions:

(1) sufficient condition: $\gamma_B(D) = \gamma_C(D)$;
(2) necessary condition: for $\forall a \in B$, a is indispensable in B to D.

Algorithm 2: Dependency with Gaussian Kernel Approximation (DGKA).

```
Input: The fuzzy relation R.
Output: The fuzzy dependency γ_B(D) of D to B.
begin
    γ_B(D) = 0;
    for each d_i ∈ U/D do
        for j = 0; j < |U|; j++ do
            find the nearest sample M_j of x_j with a different class
            γ_B(D) = γ_B(D) + sqrt(1 − R_G^2(M_j, x_j));
        end
    end
    return γ_B(D);
end
```

Theorem 2. *Given a HIS* $= \langle U, C \cup D\rangle$. $B_1 \subseteq B_2 \subseteq C$. R_1 *and* R_2 *are two* T*-equivalence relations on* U *computed with Gaussian function in* B_1 *and* B_2*, respectively. Then we have*

$$\gamma_{B_1} \leq \gamma_{B_2}. \tag{18}$$

One of the most important applications of RST is to evaluate the classification power of attributes in a decision system by computing the dependency between condition attributes and the resulting decision. The dependency function is used as a sort of heuristics in constructing efficient greedy attribute reduction algorithms [57].

In a forward greedy search, one starts with an empty set of attributes, and keeps adding features to the subset of selected attributes one by one. Each selected attribute maximizes the increment of dependence of the current subset; this implies the relevant but redundant attributes will not be included because it cannot bring much new information about classification if the attribute is redundant. Formally, Hu et al. proposed a forward search algorithm FS-GKA for feature selection based on Gaussian kernel approximation. FS-GKA algorithm does not include the process of computing the fuzzy relations. In real-world HIS, the generation of fuzzy relations will consume huge time, especially when the HIS is very large. Therefore, based on FS-GKA, we use the new Gaussian FRS approach to compute the fuzzy relation, and design a FRS Approach for Naive Feature Selection in HIS (FRSA-NFS-HIS) as follows.

Algorithm 3: A FRS Approach for Naive Feature Selection in HIS (FRSA-NFS-HIS).

```
   Input: HIS' = ⟨U', C' ∪ D', V', f'⟩ at time t + 1.
   Output: Reduction attribute set red.
1  begin
2      red = ∅, γ_red = 0;
3      while red ≠ C do
4          γ_max = 0, b = ∅;
5          for each a_i ∈ (C − red) do
6              Compute R_G^{red∪{a_i}};                 // Computing fuzzy relation by Algorithm 1.
7              γ_i = Gen_γ(R_G^{red∪{a_i}});            // Computing the dependency γ_i by DGKA algorithm.
8          end
9          find the maximal dependency γ_max and the corresponding attribute b;
10         if γ_max − γ_red > ε then
11             red = red ∪ {b}, γ_red = γ_max;
12         else
13             break;
14         end
15     end
16     return red;
17 end
```

Algorithm 3 is a non-incremental algorithm, and includes two key parts: (1) computing fuzzy relations (in Line 6); (2) computing the dependency γ_i by DGKA algorithm (in Line 7). The time complexity of part (1) is $O(m^2 \times n^2)$, and that of part (2) is $O(m^2 \times n \log n)$. Therefore, the whole time complexity of Algorithm 3 is $O(m^2 \times n^2)$, where m and n are the numbers of features and samples, respectively.

In dynamic and big information systems, Algorithm 3 is ineffective. In the next section, we will propose two FRS Algorithms for Incremental Feature Selection in HIS (FRSA-IFS-HIS) to decrease the time complexity.

Table 2
The main symbols used in the incrementally learning process.

P	The attribute set in U at time t	q_a, q_d	The added, deleted attribute at time $t+1$
$red^{(t)}$	The attribute reduct at time t	$red^{(t+1)}$	The attribute reduct at time $t+1$
$redset^{(t)}$	The reduct set at time t	$redset^{(t+1)}$	The reduct set at time $t+1$
$R_G^{(t)}$	The fuzzy relation set in U at time t	$R_G^{(t+1)}$	The fuzzy relation set in U at time $t+1$
$\gamma_{red}^{(t)}$	Approximating quality set at time t	$\gamma_{red}^{(t+1)}$	Approximating quality set at time $t+1$
$R_{red}^{(t)}$	The fuzzy relation set in reduct set at time t	$R_{red}^{(t+1)}$	The fuzzy relation set in reduct set at time $t+1$

5. A FRS approach of incremental feature selection under the variation of the attribute set

We discuss the variation of reduct in HIS when the attribute set evolves over time. Given a $HIS = \langle U, C \cup D, V, f\rangle$ at time t, $U \neq \emptyset$ and $C \cap D = \emptyset$. Suppose there are some attributes that enter into HIS or get out from HIS at time $t+1$. The fuzzy equivalence relations will be changed. And then, the attribute reduct will be changed too. Table 2 shows the main symbols used in the incrementally learning process. Let $R_G^{(t)}$ be the fuzzy equivalence relation at time t. The attribute reduct is denoted by $red^{(t)}$ at time t, and $red^{(t)} = \{red_1, red_2, ..., red_r\}$, $r = |red^{(t)}|$. $redset^{(t)} = \{rs_1, rs_2, ..., rs_r\}$, $r = |red^{(t)}|$. $rs_1 = \{red_1\}$, $rs_2 = \{red_1\} \cup \{red_2\}$ and $rs_r = \{red_1\} \cup \{red_2\} \cup ... \cup \{red_r\}$, respectively. $\gamma_{red}^{(t)}$ denotes the approximating quality of $red^{(t)}$ at time t, and $\gamma_{red}^{(t)} = \{\gamma_{red_1}, \gamma_{red_2}, ..., \gamma_{red_r}\}$. γ_{red_1} denotes the quality of rs_1, γ_{red_2} denotes the quality of rs_2 and γ_{red_r} denotes the quality of rs_r, respectively. $R_{red}^{(t)}$ denotes the fuzzy relation set of $red^{(t)}$ at time t, and $R_{red}^{(t)} = \{R_{red_1}, R_{red_2}, ..., R_{red_r}\}$. R_{red_1} denotes the fuzzy relation of rs_1, R_{red_2} denotes the fuzzy relation of rs_2 and R_{red_r} denotes the fuzzy relation of rs_r, respectively. Let $P \subseteq C$ denote the attribute set at time t and R_G^P denote the fuzzy equivalence relation under the attribute set P. Let $R_G^{(t+1)}$ be the fuzzy equivalence relation at time $t+1$, q_a denote the adding attribute and q_d be the deleting attribute at time $t+1$. Therefore, the attribute set is $P \cup \{q_a\} - \{q_d\}$ in HIS at time $t+1$. After adding and deleting attributes, the HIS will be changed into $HIS' = (U', C' \cup D', V', f')$, $P \cup \{q_a\} - \{q_d\} \subseteq C'$. The attribute reduct is denoted by $red^{(t+1)}$ at time $t+1$, and $red^{(t+1)} = \{red_1, red_2, ..., red_{r'}\}$, $r' = |red^{(t+1)}|$. $redset_{red}^{(t+1)}$ denotes the reduct set of $red^{(t+1)}$. $\gamma_{red}^{(t+1)}$ denotes the approximating quality of $red^{(t+1)}$. $R_{red}^{(t+1)}$ denotes the fuzzy relation set of $red^{(t+1)}$.

With these stipulations, we focus on the algorithms for incremental feature selection when (1) one attribute is added into the HIS at time $t+1$; (2) one attribute is deleted from the HIS at time $t+1$.

5.1. Addition of one attribute

Because Algorithm 3 includes two key parts (computing fuzzy relations and dependencies). We improve the algorithm from two steps: (1) computing fuzzy relations incrementally; (2) selecting features incrementally.

5.1.1. Computing fuzzy relations incrementally

Given a $HIS = \langle U, C \cup D, V, f\rangle$, $\forall x_i, x_k \in U$. x_i, x_k can be seen as two n-dimension vectors, and $x_i = \langle x_i^{c_1}, x_i^{c_2}, ..., x_i^{c_n}\rangle$, $x_k = \langle x_k^{c_1}, x_k^{c_2}, ..., x_k^{c_n}\rangle$, $c_j \in C$, and $j = 1, ..., n, n = |C|$. $\forall P \subseteq C$, x_i, x_k can be seen as two m-dimension vectors denoted as x_i^P and x_k^P, respectively. $x_i^P = \langle x_i^{p_1}, x_i^{p_2}, ..., x_i^{p_m}\rangle$, $x_k^P = \langle x_k^{p_1}, x_k^{p_2}, ..., x_k^{p_m}\rangle$, $p_j \in P$, and $j = 1, ..., m, m = |P|$.

If only considering one attribute a, $a \in C$, according to formula (12), the fuzzy relation between samples x_i and x_k is as follows.

$$R_G^{\{a\}}(x_i, x_k) = \exp\left(-\frac{\|x_i^{\{a\}} - x_k^{\{a\}}\|^2}{2\delta^2}\right).$$

If there are two attributes $a, b \in C$, the fuzzy relations $R_G^{\{a\}}(x_i, x_k)$, $R_G^{\{b\}}(x_i, x_k)$ between samples x_i and x_k are computed, respectively, and they can be aggregated together as shown in Proposition 3.

Proposition 3. *$\forall a, b \in C, \forall x_i, x_k \in U$, and $a \neq b, x_i \neq x_k$.*

$$R_G^{\{a\}\cup\{b\}}(x_i, x_k) = R_G^{\{a\}}(x_i, x_k) \times R_G^{\{b\}}(x_i, x_k). \tag{19}$$

Note,

$$\begin{aligned} R_G^{\{a\}\cup\{b\}}(x_i, x_k) &= \exp\left(-\frac{\|x_i^{\{a\}\cup\{b\}} - x_k^{\{a\}\cup\{b\}}\|^2}{2\delta^2}\right) = \exp\left(-\frac{d_{sa}(x_i, x_k)^2 + d_{sb}(x_i, x_k)^2}{2\delta^2}\right) \\ &= \exp\left(-\frac{\|x_i^{\{a\}} - x_k^{\{a\}}\|^2 + \|x_i^{\{b\}} - x_k^{\{b\}}\|^2}{2\delta^2}\right) = \exp\left(-\frac{\|x_i^{\{a\}} - x_k^{\{a\}}\|^2}{2\delta^2}\right) \\ &\times \exp\left(-\frac{\|x_i^{\{b\}} - x_k^{\{b\}}\|^2}{2\delta^2}\right) = R_G^{\{a\}}(x_i, x_k) \times R_G^{\{b\}}(x_i, x_k). \end{aligned}$$

If there is an attribute set $P \subset C$, and a new attribute q_a. The fuzzy relations $R_G^P(x_i, x_k)$, $R_G^{\{q_a\}}(x_i, x_k)$ between samples x_i and x_k are computed, respectively, and they can be aggregated together as Proposition 4.

Proposition 4. *Let $P \subseteq C$, $\forall q_a \in C$, and $q_a \notin Q$, $\forall x_i, x_k \in U$, and $x_i \neq x_k$.*

$$R_G^{P\cup\{q_a\}}(x_i, x_k) = R_G^P(x_i, x_k) \times R_G^{\{q_a\}}(x_i, x_k). \tag{20}$$

According to Proposition 4, it is easy to compute the fuzzy relation incrementally. Based on the feature set B and the corresponding relation R_G^B, we use Proposition 1 to redesign Algorithm 1. The time complexity to compute $R_G^{B\cup\{q_a\}}$ can be decreased to $O(n^2)$.

5.1.2. Incremental feature selection

In this step, an incrementally updating features approach on the selected features at time t is presented when a new attribute enters into the system.

Proposition 5. *Let q_a be a new attribute, $red^{(t)} = \{red_1, red_2, ..., red_r\}$ be the reduct at time t and $r = |red^{(t)}|$. $redset^{(t)} = \{red_1, red_1 \cup red_2, ..., red_1 \cup red_2 \cup ... \cup red_r\}$. The reduct at time $t+1$ can be computed as follows.*

$$red^{(t+1)} = \begin{cases} red^{(t)}, & \forall rs_i \in redset^{(t)}, \gamma_{rs_{i-1}\cup\{q_a\}} \leq \gamma_{rs_i} \text{ and } \gamma_{rs_r\cup\{q_a\}} - \gamma_{rs_r} \leq \varepsilon \\ red^{(t)} \cup X, & \forall rs_i \in redset^{(t)}, \gamma_{rs_{i-1}\cup\{q_a\}} \leq \gamma_{rs_i} \text{ and } \gamma_{rs_r\cup\{q_a\}} - \gamma_{rs_r} > \varepsilon \\ rs_{min-1} \cup \{q_a\} \cup Y, & \exists rs_i \in redset^{(t)}, \gamma_{rs_{i-1}\cup\{q_a\}} > \gamma_{rs_i}, \end{cases} \tag{21}$$

where,

- *X is the attribute set which should be added into $red^{(t+1)}$, and $X \subseteq C - red^{(t)}$.*
- *$rs_{min}^{(t)}$ is the most short subset in $redset^{(t)}$ which satisfies $\gamma_{rs_{i-1}\cup\{q_a\}} > \gamma_{rs_i}$.*
- *Y is the attribute set which should be added into $red^{(t+1)}$, and $Y \subseteq C - rs_{min-1}$.*

According to Proposition 5, when the new attribute is added, $red^{(t+1)} = \{red_1, ..., red_{Ord}, ..., red_r\}$, Ord denotes the place (or order) of the new attribute in $red^{(t+1)}$. There are three cases:

(1) $1 \leq Ord \leq r$, the new attribute is in $red^{(t+1)}$, and its place is among the places of the elements in $red^{(t)}$;
(2) $Ord = r + 1$, the new attribute is in $red^{(t+1)}$, and its place is at the tail of $red^{(t)}$;
(3) $Ord > r + 1$, the new attribute is not in $red^{(t+1)}$.

Algorithm 4 is an incremental algorithm for feature selection while one attribute is added into the HIS. In Line 2, we compute the fuzzy relation of the new attribute q_a. The time complexity is $O(n^2)$. Lines 3–11 locate the new attribute's place Ord in the reduct $red^{(t)}$. If the new attribute is added into $red^{(t+1)}$, Line 7 finishes the adding task, and in Line 8, Algorithm 5 is called to incrementally select the following attributes which will be in the reduct. Let $i = Ord$. The

Algorithm 4: A FRS Approach for Incremental Feature Selection in HIS under one Attribute being Added (FRSA-IFS-HIS(AA)).

Input:
(1) $HIS = (U, C \cup D, V, f)$ at time t;
(2) New attribute q_a;
(3) Reduction attribute set $red^{(t)}$, and its corresponding $redset^{(t)}$, $\gamma_{rs}^{(t)}$ and $R_{rs}^{(t)}$;
Output: $red^{(t+1)}$, $redset^{(t+1)}$, $\gamma_{rs}^{(t+1)}$ and $R_{rs}^{(t+1)}$.
1 **begin**
2 Compute $R_G^{\{q_a\}}$;
3 **for** $i = 1; i \le r; i++$ **do**
4 $S = R_{rs_{i-1}} \times R_G^{\{q_a\}}$; `// Incremental generate fuzzy relation of` $rs_i \cup \{q_a\}$
5 $\gamma_{rs_{i-1} \cup \{q_a\}} = \mathrm{Gen\_}\gamma(S)$; `// Compute the dependency` $\gamma_{rs_{i-1} \cup \{q_a\}}$ `of` D `to` $red \cup \{q_a\}$`, time complexity // is` $\mathrm{O}(n \log n)$
6 **if** $\gamma_{rs_{i-1} \cup \{q_a\}} > \gamma_{rs_i}$ **then**
7 $red = rs_{i-1} \cup \{q_a\}$, $rs_i = red$, $\gamma_{rs_i} = \gamma_{rs_{i-1} \cup \{q_a\}}$, $R_{rs_i} = S$;
8 Gen_Red_Incre(C,rs_i,γ_{rs_i},R_{rs_i}); `// time complexity is` $\mathrm{O}((m-i)^2 \times n \log n)$.
9 **break**;
10 **end**
11 **end**
12 **if** $i == r + 1$ **then** `// It means that` $\{q_a\}$ `has not been added into reduct`
13 $S = R_{rs_r} \times R_G^{\{a_i\}}$;
14 $\gamma_{rs_r \cup \{q_a\}} = \mathrm{Gen\_}\gamma(S)$;
15 **if** $\gamma_{rs_r \cup \{q_a\}} - \gamma_{rs_r} > \varepsilon$ **then** `//` $\{q_a\}$ `needs to be added into reduct`
16 $red = rs_r \cup \{q_a\}$, $rs_{r+1} = red$, $\gamma_{rs_{r+1}} = \gamma_{rs_r \cup \{q_a\}}$, $R_{rs_{r+1}} = S$;
17 Gen_Red_Incre(C,rs_{r+1},$\gamma_{rs_{r+1}}$,$R_{rs_{r+1}}$);
18 **end**
19 **end**
20 **return** red, $redset$, γ_{rs}, R_{rs}.
21 **end**

time complexity is $\mathrm{O}(i \times n \log n + (m-i)^2 \times n \log n) = \mathrm{O}(i \times n \log n + (m-i)^2 \times n \log n)$. The worst case is the new attribute is at the head of $red^{(t+1)}$ ($Ord = 1$), which means re-doing feature selection. The time complexity is $\mathrm{O}(m^2 \times n \log n)$. Once $Ord > r$, Lines 12–18 judge whether the new attribute should be added to the tail of $red^{(t)}$. The best case is $Ord > r + 1$, the new attribute will not be added into $red^{(t+1)}$. The algorithm will not call Algorithm 5, and there are only $r + 1$ steps to compute dependencies. The time complexity is $\mathrm{O}(r \times n \log n)$. Therefore, the worst case time complexity is $\mathrm{O}(n^2 + m^2 \times n \log n)$, and the best case time complexity is $\mathrm{O}(n^2 + |red| \times n \log n)$.

Algorithm 5: Gen_Red_Incre(C,rs_{r+1},$\gamma_{rs_{r+1}}$,$R_{rs_{r+1}}$).

Input: C, rs_i, γ_{rs_i}, R_{rs_i}
Output: red, $redset$, γ_{rs}, R_{rs}
begin
$red = rs_i$, $\gamma_{red} = \gamma_{rs_i}$, $R_{red} = R_{rs_i}$, $j = 0$;
while $red \neq C$ **do**
$R_{max}[,] = 1$, $\gamma_{max} = 0$, $b = \emptyset$;
for *each* $a_i \in (C - red)$ **do**
$S = R_{rs_i} \times R_G^{\{a_i\}}$; `// Because` R_{rs_i} `and` $R_G^{\{a_i\}}$ `have been generated, they will not be computed // again.`
$\gamma_i = Gen\_\gamma(S)$; `// Computing the dependency` γ_i `of` D `to` $red \cup \{a_i\}$
end
find the maximal dependency γ_{max} and the corresponding attribute b and fuzzy relation R_{max};
if $\gamma_{max} - \gamma_{red} > \varepsilon$ **then**
$red = red \cup \{b\}$, $\gamma_{red} = \gamma_{max}$, $R_{red} = R_{max}$, $j = j + 1$;
$rs_{i+j} = red$, $\gamma_{rs_{i+j}} = \gamma_{red}$, $R_{rs_{i+j}} = R_{red}$;
else
break;
end
end
return red, $redset$, γ_{rs}, R_{rs};
end

Comparing with Algorithm 3, Algorithm 5 does not include the process of computing the fuzzy relation. Therefore, its time complexity is $O(|C - rs_i|^2 \times n \log n)$, where rs_i represents the unchanged part of the reduct.

5.2. Deletion of one attribute

Before the attribute is deleted, the fuzzy relations have been computed. Therefore, we only improve the algorithm in selecting features incrementally.

Proposition 6. *Let q_d be an attribute to be deleted, and $red^{(t)}$ be the reduct at time t. The reduct at time $t+1$ can be computed as follows.*

$$red^{(t+1)} = \begin{cases} red^{(t)}, & q_d \notin red^{(t)} \\ red^{(t)}_{max} \cup Z, & otherwise, \end{cases} \tag{22}$$

where,

- *$red^{(t)}_{max}$ is the longest subset in $redset^{(t)}$ which does not include q_d.*
- *Z is the attribute set which should be added into $red^{(t+1)}$, and $Z \subseteq C - \{q_d\} - red^{(t)}_{max}$.*

According to Proposition 6, when an attribute is deleted, there are two cases:

(1) The deleting attribute is not in $red^{(t)}$;
(2) The deleting attribute is in $red^{(t)}$.

Let Ord denote the place (or order) of the deleting attribute in $red^{(t)}$. In case (1), $Ord > r$; In case (2), $1 \leq Ord \leq r$.

Algorithm 6: A Fuzzy Rough Sets Approach for Incremental Feature Selection in HIS under Attribute being Deleted (FRSA-IFS-HIS(AD)).

Input:
(1) $HIS = (U, C \cup D, V, f)$ at time t;
(2) Reduction attribute set $red^{(t)}$, and its corresponding $redset^{(t)}$, $\gamma^{(t)}_{rs}$ and $R^{(t)}_{rs}$;
(3) The deleted attribute q_d.
Output: $red^{(t+1)}$, $redset^{(t+1)}$, $\gamma^{(t+1)}_{rs}$ and $R^{(t+1)}_{rs}$.

1 **begin**
2   **for** ($i = r; i \geq 1$ *and* $q_d \in rs_i; i - -$) **do**
3   **end**
4   **if** $i > 0$ **then**   // It means that $\{q_d\}$ is in the reduct
5   Gen_Red_Incre(C,rs_{i-1},$\gamma_{rs_{i-1}}$,$R_{rs_{i-1}}$);
6   **else**
7         $red = red^{(t)}$, $redset = redset^{(t)}$, $\gamma_{rs} = \gamma^{(t)}_{rs}$, $R_{rs} = R^{(t)}_{rs}$;
8   **end**
9   **return** red, $redset$, γ_{rs}, R_{rs};
10 **end**

Algorithm 6 is an incremental algorithm for feature selection while one attribute is deleted from HIS. Lines 2–6 locate the deleting attribute's place Ord in $red^{(t)}$. If $1 \leq Ord \leq r$, Line 7 finishes the deleting task, and Line 5 calls Algorithm 5 to incrementally select the features which place are after the deleting attribute. The worst case is $Ord = 1$. The time complexity is $O(m^2 \times n \log n)$. Once $Ord > r$, the reduct will not be changed. The time complexity is $O(r)$. Therefore, the worst case time complexity is $O(m^2 \times n \log n)$, and the best case time complexity is $O(r)$.

6. Experimental evaluation

We compare the computational time of the non-incremental algorithm (FRSA-NIFS-HIS) and the two incremental algorithms (FRSA-IFS-HIS(AA) and FRSA-IFS-HIS(AD)) on different datasets. To simplify the expression, let FRSA-IFS-HIS denote the two incremental algorithms.

Table 3
Descriptions of datasets.

ID	Data set	Samples	Features	Feature types	Decision class
1	Wine	178	18	Boolean, Real	3
2	Heparits	155	19	Categorical, Boolean, Real	2
3	Horse	368	26	Categorical, Boolean, Real	2
4	Iono	351	34	Boolean, Real	2
5	Credit	690	15	Categorical, Boolean, Real	2
6	Credit(set)	690	16	Categorical, Boolean, Real, Set	2
7	German	1000	20	Categorical, Boolean, Real	2
8	Sick	2800	28	Categorical, Boolean, Real	2
9	Bands	540	40	Categorical, Boolean, Real	2

Table 4
Number of selected features with FRSA-IFS-HIS, Gaussian kernel approximation and fuzzy entropy.

ID	Data set	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
1	Wine	5	6	6
2	Heparits	3	7	4
3	Horse	10	4	5
4	Iono	8	7	9
5	Credit	5	5	2
6	Credit(set)	6	–	–
7	German	7	11	4
8	Sick	7	6	6
9	Bands	12	11	10

Experiments are performed on a 2.2 GHz, 8 cores and 16 threads Server with 8 GB of memory, running Win 7. Algorithms are coded in SQL Server 2008. Eight datasets (including hybrid features) used here come from the UCI Machine Learning Repository (http://www.ics.uci.edu/) (see Table 3). To evaluate the capability of dealing with the set-valued attributes, we compose a new datasets Credit (set) based on Credit. In Credit, three boolean attributes 1, 9, 10 are combined to a new set-valued attribute 16 (shown as Credit (set) in Table 3). In experiments, learning algorithms such as CART, Poly SVM and RBF SVM are used. The experiments are run in a 10-fold cross validation mode. The parameters of the poly SVM and RBF SVM are taken as the default values (the use of the weka 3.6.9 toolkit smo svm). The discretization method is the unsupervised attributes discretization.

In computing the fuzzy relations with Gaussian kernel, one should specify kernel parameter δ. In [22], Hu et al. gave an approach of setting δ. Therefore, according to the approach, we taken the values of δ from $[0.1, 0.36]$ randomly.

6.1. Comparisons of classification performance between FRSA-IFS-HIS and others

Firstly, we evaluate the effectiveness of the attribute reduction with three functions: FRSA-IFS-HIS, Gaussian kernel approximation (Gaussian) in [22] and fuzzy entropy in Gaussian kernel approximation (Entropy). Table 4 shows the number of selected features by the three feature selection techniques in nine datasets.

From Table 4, we can get the following results: First, like other feature selection techniques, algorithm FRSA-IFS-HIS can delete most of unnecessary attributes. Second, algorithm FRSA-IFS-HIS is effective in credit (set), but the other two techniques are not.

Although the algorithm FRSA-IFS-HIS can select features in all datasets, the classification performance of the selected features have to be tested. There are several criteria that can be used to test the performance of the proposed methods. For example, classification accuracy, sensitivity, specificity and the area under the ROC curve (AUC). In this paper, we use classification accuracy and AUC to test the classification performance. We build classification models with the selected features and test their classification performance based on 10-fold cross validation. We compare the raw data, algorithm FRSA-IFS-HIS, Gaussian kernel based FRS in [22], Fuzzy entropy in Tables 5–10, where learning algorithms CART, RBF SVM and Poly SVM are introduced to evaluate the selected features. In Tables 5, 6, 7, we

Table 5
Classification accuracies based on CART (%).

Data set	Raw data	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
Wine	89.89	**91.57**	89.89	89.89
Heparits	79.35	76.77	76.77	**81.94**
Horse	85.05	**85.3**	67.3	69.67
Iono	90.03	**91.74**	91.45	89.46
Credit	84.35	**85.51**	**85.51**	**85.51**
German	72.1	73.1	72.1	**75**
Sick	97.39	**97.39**	97.21	96.85
Bands	60	**60**	60	60
Credit(set)	84.35	**85.51**	–	–

Table 6
Classification accuracies based on RBF SVM (%).

Data set	Raw data	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
Wine	94.38	**96.63**	**96.63**	**96.63**
Heparits	79.35	**79.35**	**79.35**	**79.35**
Horse	83.15	**79.62**	63.67	63.67
Iono	90.31	84.90	84.05	**89.74**
Credit	85.51	**85.51**	**85.51**	**85.51**
German	73.3	74.9	**75.1**	73.6
Sick	96.89	**97.1**	96.82	96.93
Bands	72.78	**72.78**	72.78	72.78
Credit(set)	84.35	**85.51**	–	–

Table 7
Classification accuracies based on Poly SVM (%).

Data set	Raw data	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
Wine	96.07	**95.51**	**95.51**	**95.51**
Heparits	81.94	**82.58**	**82.58**	80.65
Horse	79.62	**82.07**	62.77	65.67
Iono	91.45	**91.17**	**91.17**	90.6
Credit	85.36	84.93	84.93	**85.51**
German	71.7	**74.7**	73.2	71
Sick	97.1	97.2	**97.25**	95.82
Bands	77.2	**77.2**	77.2	77.2
Credit(set)	84.35	**85.51**	–	–

report the classification accuracies of the selected features based on the use of the Cart, RBF SVM and Poly SVM. The higher the classification accuracy is, the more superior the classification performance we obtain. In Tables 8, 9, 10, we report the area under the ROC curve of the selected features. The larger the AUC is, the more superior the classification performance we get.

From Tables 5–10, it is clear that the classification performance by the results of algorithm FRSA-IFS-HIS, Gaussian kernel based FRS and Fuzzy entropy is better than that by the raw data. It is also concluded that, in most datasets, algorithm FRSA-IFS-HIS performs better than the other two feature selection techniques.

6.2. *Comparisons of the computational time between FRSA-NFS-HIS and FRSA-IFS-HIS*

Generally, we perform experimental analysis with applying the non-incremental algorithm FRSA-NFS-HIS along with our proposed incremental algorithms when the one attribute is added or deleted from the HIS, respectively. From Algorithms 4 and 6, we know the place (*Ord*) of the adding (deleting) attribute which will be added into (deleted from) the reduct will effect the executing time of the algorithms. In this paper, we will analyze the relations between *Ord* and the executing time in FRSA-NFS-HIS and FRSA-IFS-HIS, respectively.

Table 8
The area under the ROC curve (AUC) based on CART (%).

Data set	Raw data	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
Wine	93.6	**94.33**	92.5	92.5
Heparits	64.8	64.4	73.3	**75**
Horse	83.8	**82.1**	65.4	69.8
Iono	87.6	**90.9**	89.9	88.9
Credit	86.3	**83.7**	**83.7**	**83.7**
German	68.7	70.8	69.1	**72.1**
Sick	81.3	83	**84.1**	81.9
Bands	55.2	**55.2**	54.8	54.9
Credit(set)	86.3	**83.7**	–	–

Table 9
The area under the ROC curve (AUC) based on RBF SVM (%).

Data set	Raw data	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
Wine	97.5	**97.3**	**97.3**	**97.3**
Heparits	50	50	50	50
Horse	81	**78.6**	50	50
Iono	87.4	**93.2**	78.1	86.2
Credit	86.2	**86.2**	**86.2**	**86.2**
German	60.9	**65.4**	63.8	61.8
Sick	82.8	**85.6**	81.9	82.8
Bands	69.7	**70**	69.7	69.7
Credit(set)	86.2	**86.2**	–	–

Table 10
The area under the ROC curve (AUC) based on Poly SVM (%).

Data set	Raw data	FRSA-IFS-HIS	Gaussian	Fuzzy entropy
Wine	97.7	**96.5**	96.1	96.1
Heparits	67.8	65.9	**70.2**	50
Horse	78.9	**80**	62.9	57.9
Iono	89.8	**83.1**	89.8	89
Credit	86	85.6	85.6	**86.2**
German	65.5	62.3	**65.3**	52.5
Sick	84.2	**85.4**	**85.4**	78.9
Bands	76.66	**75.4**	73.4	72.3
Credit(set)	86	**85.6**	–	–

6.2.1. Comparisons between FRSA-NFS-HIS and FRSA-IFS-HIS(AA)

We compare the efficiency of non-incremental (FRSA-NFS-HIS) and incremental (FRSA-IFS-HIS(AA)) algorithms for selecting features when adding a new attribute into the datasets. On the nine datasets, the experimental results are presented in Fig. 1.

In each sub-figure of Fig. 1, the x-coordinate pertains to the new attribute's place (*Ord*) which will be added into the reduct, while the y-coordinate concerns the computational time. Fig. 1 displays more detailed change trend of each of the two algorithms with the increasing of the *Ord*.

From Fig. 1, the elapsed time of FRSA-IFS-HIS(AA) is much smaller than RSA-NFS-HIS. With the increasing of the *Ord*, the elapsed time is relatively stable in the non-incremental method (FRSA-NFS-HIS), but the elapsed time decreases and tends to a stable value in the incremental updating method (FRSA-IFS-HIS(AA)). From Table 4, most of attributes will not be included in reduct. If the adding attribute will not be included in reduct, the computing time of the algorithm FRSA-IFS-HIS(AA) is very low. In general, FRSA-IFS-HIS(AA) is much faster than FRSA-NFS-HIS when one attribute is added into HIS.

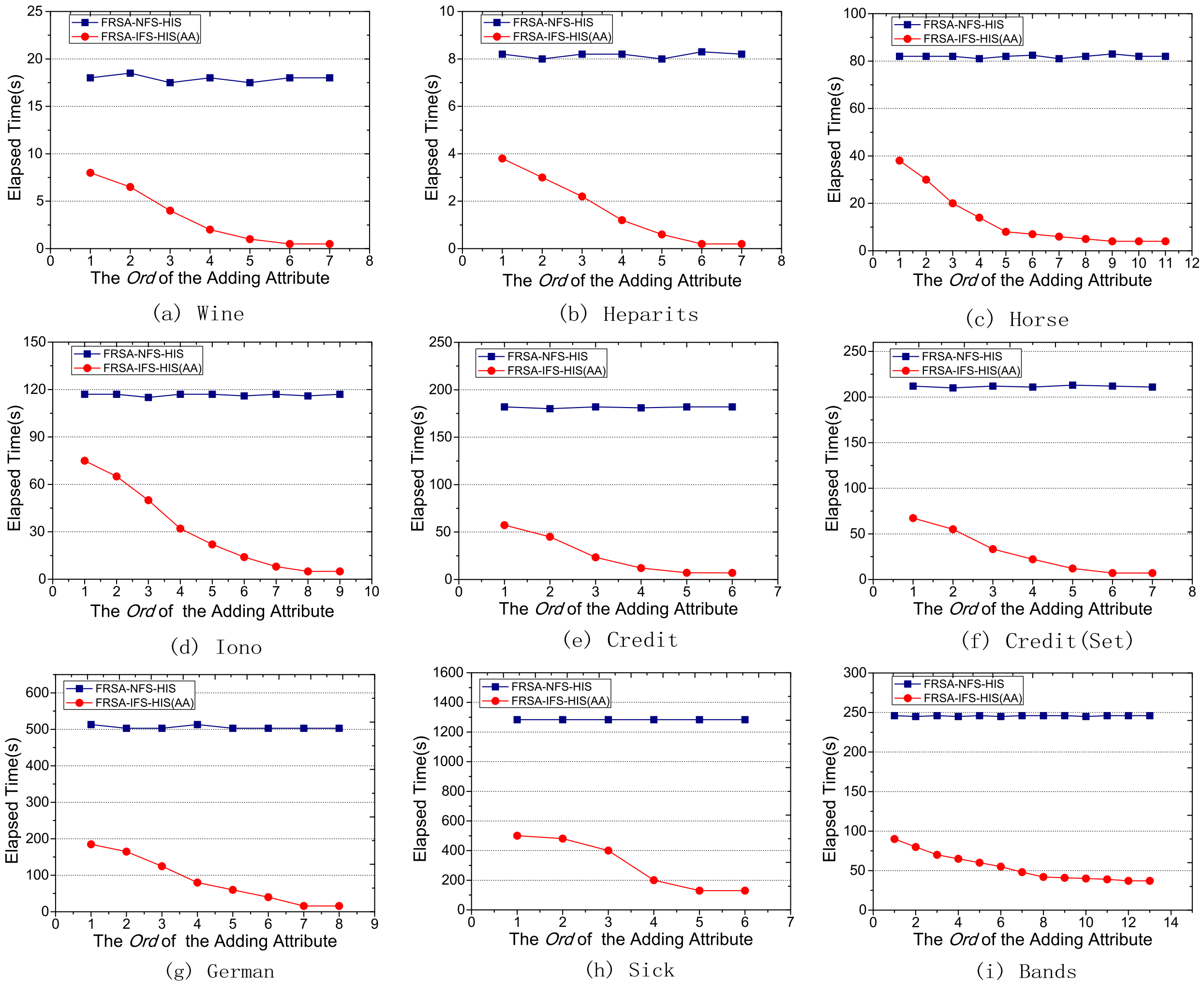

Fig. 1. Elapsed time of FRSA-NFS-HIS and FRSA-IFS-HIS(AA) versus the *Ord* of the adding attribute.

6.2.2. *Comparisons between FRSA-NFS-HIS and FRSA-IFS-HIS(AD)*

We compare the efficiency of non-incremental (FRSA-NFS-HIS) and incremental (FRSA-IFS-HIS(AD)) algorithms for selecting features when deleting an attribute from the datasets. On the nine datasets, the experimental results are presented in Fig. 2.

In each sub-figure of Fig. 2, the x-coordinate pertains to the deleting attribute's place (*Ord*) in the reduct, while the y-coordinate concerns the computational time. Fig. 2 displays more detailed change trend of each of the two algorithms with the increasing of the *Ord*.

From Fig. 2, the elapsed time of FRSA-IFS-HIS(AD) is much smaller than that of RSA-NFS-HIS. With the increasing of the *Ord*, the elapsed time is relatively stable in the non-incremental method (FRSA-NFS-HIS), but the elapsed time decreases and tends to 0 in the incremental updating method (FRSA-IFS-HIS(AD)). If *Ord* is larger than the length of the reduct, it is clear that the incremental computational time is close to 0. Therefore, in general, FRSA-IFS-HIS(AA) is much faster than FRSA-NFS-HIS when one attribute is deleted from HIS.

6.2.3. *Comparison of computational time between FRSA-NFS-HIS and FRSA-IFS-HIS in large data sets*

To analyze the efficiency of FRSA-NFS-HIS and FRSA-IFS-HIS on big data, we amplify the nine datasets in Table 3 by 1, 2, 4, 8, 16 and 32 times, respectively and use them in the experiments for the comparison of the two

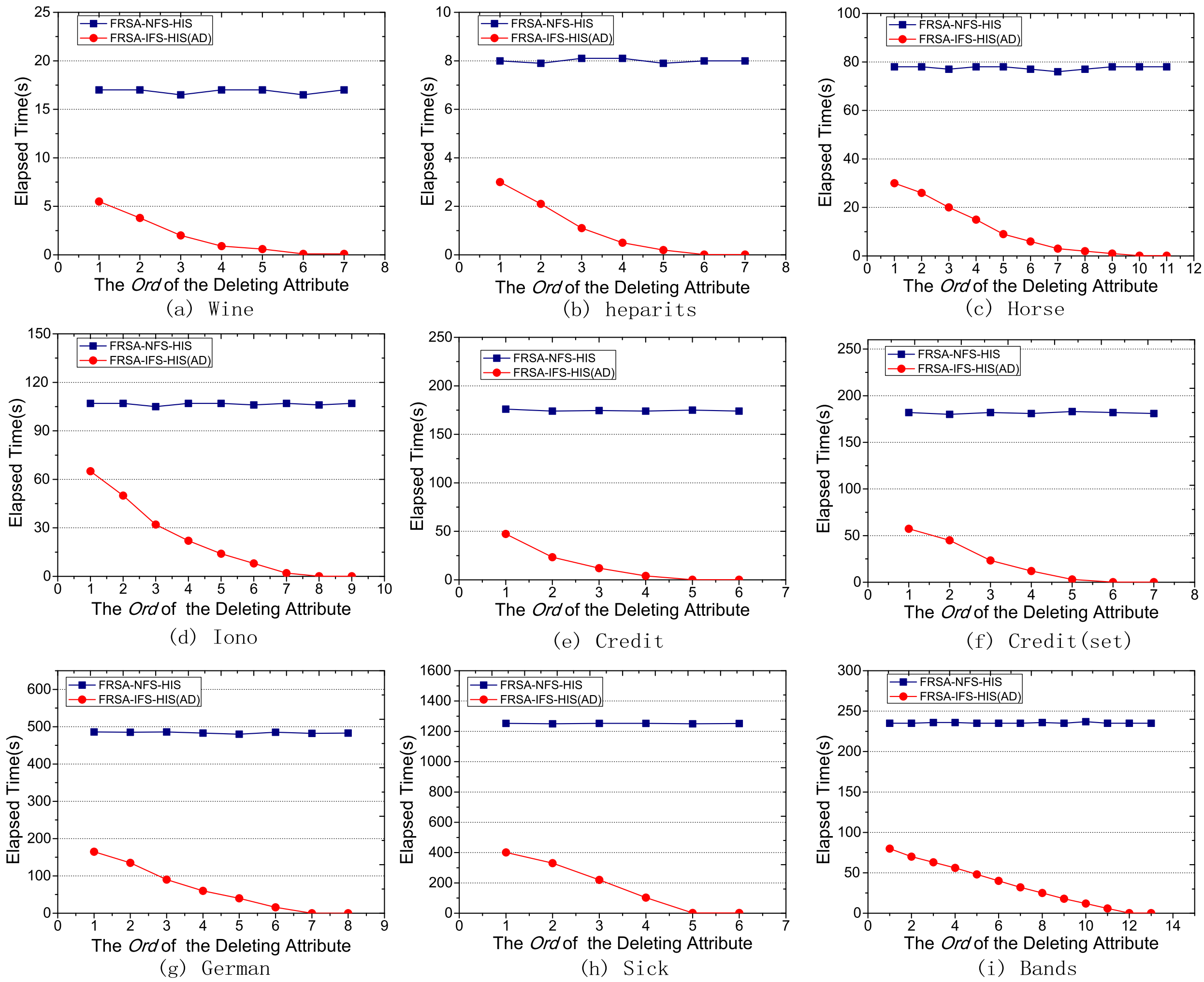

Fig. 2. Elapsed time of FRSA-NFS-HIS and FRSA-IFS-HIS(AD) versus the *Ord* of the deleting attribute.

algorithms. When adding or deleting an attribute from the extended datasets, the experimental results are presented in Fig. 3.

In each sub-figure of Fig. 3, the x-coordinate pertains to the magnification times of the datasets, while the y-coordinate concerns the average computation time where one attribute is added or deleted. Fig. 3 displays more detailed change trend of the two algorithms with the increasing of the magnification times.

From Fig. 3, with the increasing of the magnification times, the elapsed time of both the two algorithms increase, but the elapsed time of FRSA-IFS-HIS is much smaller than that of RSA-NFS-HIS. Therefore, in general, the incremental algorithm FRSA-IFS-HIS can get better efficiency than FRSA-NFS-HIS when one attribute is added or deleted from HIS in big data.

7. Conclusions

FRS is an effective approach for feature selection in HIS. The incremental technique is an effective way to maintain knowledge in the dynamic environment. These two techniques are combined to maintain knowledge in dynamic and hybrid environment. In this paper, we proposed a novel HD distance that can deal with different types of data (*e.g.*, boolean, categorical, real-valued and set-valued). And then HD distance was applied into Gaussian kernel with FRS. We discussed the principles of updating features when a new feature is added or an old one is deleted, and then proposed the corresponding incremental algorithms for feature selection in HIS. Experimental studies on eight UCI

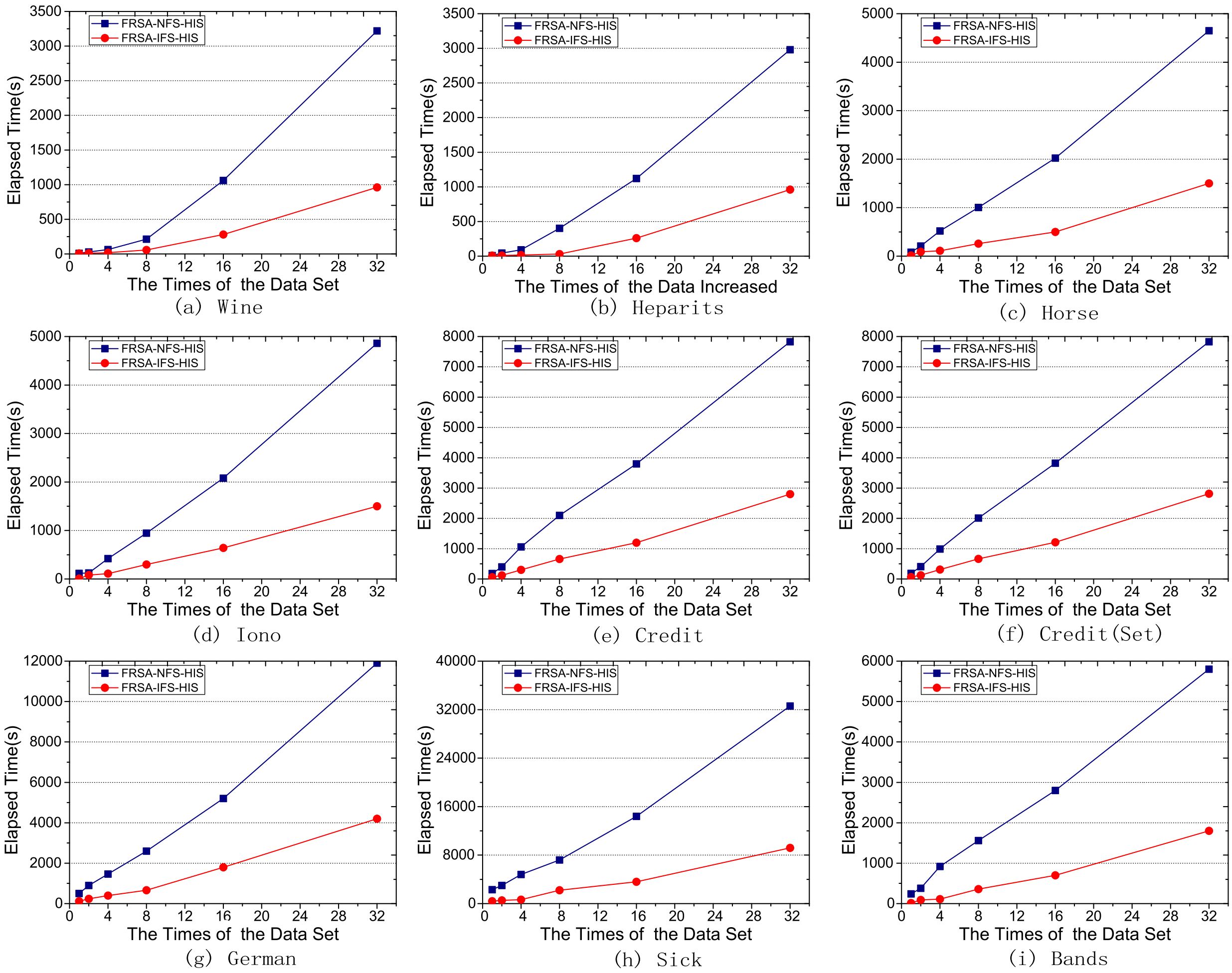

Fig. 3. Average elapsed time of FRSA-NFS-HIS and FRSA-IFS-HIS versus the magnification times of the datasets.

data sets and an artificial data set showed that the incremental algorithms can improve the computational efficiency for updating the feature set when the information system varies over time. It is validated that the proposed approaches can be applied effectively for feature selection in big data. An information system consists of objects, attributes, and the domain of attributes, all of that may change over time. In the future, the variation of multi-attributes and the objects in HIS will be taken into consideration in terms of incremental updating knowledge.

Acknowledgement

This work is supported by the National Science Foundation of China (Nos. 61175047, 61100117) and NSAF (No. U1230117), the Youth Social Science Foundation of the Chinese Education Commission (No. 11YJC630127), the Scientific Research Fundation of Sichuan Provincial Education Department (No. 13ZB0210), the Fundamental Research Funds for the Central Universities (Nos. SWJTU11ZT08, SWJTU12CX117, SWJTU12CX091), and the Doctoral Innovation Funds of Southwest Jiaotong University (No. 2013ZAP).

References

[1] N. Kwak, C.H. Choi, Input feature selection for classification problems, IEEE Trans. Neural Netw. 13 (2002) 143–159.
[2] D.P. Muni, N.R. Das Pal, Genetic programming for simultaneous feature selection and classifier design, IEEE Trans. Syst. Man Cybern., Part B 36 (1) (2006) 106–117.

[3] C.L. Huang, C.J. Wang, A GA-based feature selection and parameters optimization for support vector machines, Expert Syst. Appl. 31 (2006) 231–240.
[4] C.L. Huang, J.F. Dun, A distributed PSO-SVM hybrid system with feature selection and parameter optimization, Appl. Soft Comput. 8 (2008) 1381–1391.
[5] S.-W. Lin, S.-C. Chen, W.-J. Wu, C.-H. Chen, Parameter determination and feature selection for back-propagation network by particle swarm optimization, Knowl. Inf. Syst. 21 (2) (2009) 249–266.
[6] S.M. Vieira, J.M.C. Sousa, T.A. Runkler, Two cooperative ant colonies for feature selection using fuzzy models, Expert Syst. Appl. 37 (2010) 2714–2723.
[7] S.M. Vieira, L.F. Mendonc, G.J. Farinha, J.M.C. Sousa, Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients, Appl. Soft Comput. 13 (2013) 3494–3504.
[8] X.Y. Wang, J. Yang, X.L. Teng, W.J. Xia, R. Jensen, Feature selection based on rough sets and particle swarm optimization, Pattern Recognit. Lett. 28 (4) (2007) 459–471.
[9] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[10] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[11] X.P. Yang, Y. Yang, Independence of axiom sets on intuitionistic fuzzy rough approximation operators, Int. J. Mach. Learn. Cybern. 4 (5) (2013) 505–513.
[12] W.H. Xu, W.X. Sun, Y.F. Liu, W.X. Zhang, Fuzzy rough set models over two universes, Int. J. Mach. Learn. Cybern. 4 (6) (2013) 631–645.
[13] Z.T. Gong, X.X. Zhang, Variable precision intuitionistic fuzzy rough sets model and its application, Int. J. Mach. Learn. Cybern. 5 (2) (2014) 263–280.
[14] X.P. Wang, E.C.C. Tsang, S.Y. Zhao, D.G. Chen, D.S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set techniques, Inf. Sci. 177 (2007) 4493–4514.
[15] X.Z. Wang, J.H. Zhai, S.X. Lu, Induction of multiple fuzzy decision trees based on rough set technique, Inf. Sci. 178 (16) (2008) 3188–3202.
[16] I.K. Ludmila, Fuzzy rough sets: application to feature selection, Fuzzy Sets Syst. 51 (2) (1992) 147–153.
[17] Y.H. Qian, J.Y. Liang, W. Wei, Consistency-preserving attribute reduction in fuzzy rough set framework, Int. J. Mach. Learn. Cybern. (2012) 287–299.
[18] R. Jensen, Q. Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets Syst. 141 (3) (2004) 469–485.
[19] D.G. Chen, S.Y. Zhao, Local reduction of decision system with fuzzy rough sets, Fuzzy Sets Syst. 161 (2010) 1871–1883.
[20] M. Pradipta, G. Partha, Fuzzy-rough simultaneous attribute selection and feature extraction algorithm, IEEE Trans. Cybern. 43 (4) (2013) 1166–1177.
[21] Q.H. Hu, D.R. Yu, W. Pedrycz, D.G. Chen, Kernelized fuzzy rough sets and their applications, IEEE Trans. Knowl. Data Eng. 23 (11) (2011) 1649–1667.
[22] Q.H. Hu, L. Zhang, D.G. Chen, W. Pedrycz, D.R. Yu, Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications, Int. J. Approx. Reason. 51 (2010) 453–471.
[23] J. Camacho, Visualizing big data with compressed Score Plots: approach and research challenges, Chemom. Intell. Lab. Syst. 135 (2014) 110–125.
[24] K. Ackermann, S.D. Angus, A resource efficient big data analysis method for the social sciences: the case of global IP activity, Proc. Comput. Sci. 29 (2014) 2360–2369.
[25] C. Dobrea, F. Xhafa, Intelligent services for big data science, Future Gener. Comput. Syst. 37 (2014) 267–281.
[26] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, P. Tufano, Analytics: the real world use of big data, IBM Institute for Business Value—Executive report, IBM Institute for Business Value, 2012.
[27] F. Altiparmak, E. Tuncel, H. Ferhatosmanoglu, Incremental maintenance of online summaries over multiple streams, IEEE Trans. Knowl. Data Eng. 20 (2) (2008) 216–229.
[28] T.C. Chou, M.C. Chen, Using incremental PLSI for threshold resilient online event analysis, IEEE Trans. Knowl. Data Eng. 20 (3) (2008) 289–299.
[29] J.B. Zhang, T.R. Li, D. Ruan, Z.Z. Gao, C.B. Zhao, A parallel method for computing rough set approximations, Inf. Sci. 194 (2012) 209–223.
[30] J.B. Zhang, T.R. Li, Y. Pan, Parallel rough set based knowledge acquisition using MapReduce from big data, in: ACM SIGKDD12 Big Data Mining (BigMine'12), Workshop, China, Beijing, 2012, pp. 20–27.
[31] L. Shan, W. Ziarko, Data-based acquisition and incremental modification of classification rules, Comput. Intell. 11 (1995) 357–370.
[32] W. Bang, Z. Bien, New incremental learning algorithm in the framework of rough set theory, Int. J. Fuzzy Syst. 1 (1999) 25–36.
[33] D. Liu, T.R. Li, D. Ruan, J.B. Zhang, Incremental learning optimization on knowledge discovery in dynamic business intelligent systems, J. Glob. Optim. 51 (2011) 325–344.
[34] Z. Zheng, G.Y. Wang, A rough set and rule tree based incremental knowledge acquisition algorithm, Fundam. Inform. 59 (2–3) (2004) 299–313.
[35] L. Wang, Y. Wu, G.Y. Wang, An incremental rule acquisition algorithm based on variable precision rough set model, J. Chongqing Univ. Posts Telecommun. Nat. Sci. 17 (6) (2005) 709–713.
[36] J.B. Zhang, T.R. Li, D. Ruan, D. Liu, Neighborhood rough sets for dynamic data mining, Int. J. Intell. Syst. 27 (2012) 317–342.
[37] S.Y. Li, T.R. Li, D. Liu, Dynamic maintenance of approximations in dominance-based rough set approach under the variation of the object set, Int. J. Intell. Syst. 28 (8) (2013) 729–751.
[38] C. Luo, T.R. Li, H.M. Chen, D. Liu, Incremental approaches for updating approximations in set-valued ordered information systems, Knowl.-Based Syst. 50 (2013) 218–233.
[39] J.B. Zhang, T.R. Li, H.M. Chen, Composite rough sets for dynamic data mining, Inf. Sci. 257 (2014) 81–100.
[40] A.P. Zeng, T.R. Li, C. Luo, An incremental approach for updating approximations of gaussian kernelized fuzzy rough sets under the variation of the object set, Comput. Sci. 40 (7) (2013) 20–27 (in Chinese).

[41] F. Hu, G. Wang, H. Huang, Y. Wu, Incremental attribute reduction based on elementary sets, Proceedings of 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, in: LNAI, vol. 3641, Regina, Canada, 2005, pp. 185–193.
[42] Z.T. Liu, An incremental arithmetic for the smallest reduction of attributes, Acta Electron. Sin. 27 (11) (1999) 96–98.
[43] M. Yang, An incremental updating algorithm for attributes reduction based on the improved discernibility matrix, Chinese J. Comput. 30 (5) (2007) 815–822.
[44] J.Y. Liang, F. Wang, C.Y. Dang, Y.H. Qian, A group incremental approach to feature selection applying rough set technique, IEEE Trans. Knowl. Data Eng. 26 (2) (2014) 294–308.
[45] C.C. Chan, A rough set approach to attribute generalization in data mining, Inf. Sci. 107 (1998) 177–194.
[46] T.R. Li, Y. Xu, A generalized rough set approach to attribute generalization in data mining, J. Southwest Jiaotong Univ. (Engl. Ed.) 8 (1) (2000) 69–75.
[47] T.R. Li, D. Ruan, G. Wets, J. Song, Y. Xu, A rough sets based characteristic relation approach for dynamic attribute generalization in data mining, Knowl.-Based Syst. 20 (5) (2007) 485–494.
[48] Y. Cheng, The incremental method for fast computing the rough fuzzy approximations, Data Knowl. Eng. 70 (2011) 84–100.
[49] J. Zhang, T. Li, D. Liu, An approach for incremental updating approximations in variable precision rough sets while attribute generalizing, in: Proceedings of 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, Hangzhou, China, 2010, pp. 77–81.
[50] S.Y. Li, T.R. Li, D. Liu, Incremental updating approximations in dominance-based rough sets approach under the variation of the attribute set, Knowl.-Based Syst. 40 (2013) 17–26.
[51] C. Luo, T.R. Li, H.M. Chen, Dynamic maintenance of approximations in set-valued ordered decision systems under the attribute generalization, Inf. Sci. 257 (2014) 210–228.
[52] Y.Y. Yao, Combination of rough and fuzzy sets based on α-level sets, in: T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining: Analysis for Imprecise Data, Kluwer Academic Publishers, Boston, 1997, pp. 301–321.
[53] S.L. Chen, Fuzzy Theory and Its Applications, Science Press, Beijing, 2006, pp. 10–123.
[54] N.N. Morsi, M.M. Yakout, Axiomatics for fuzzy rough set, Fuzzy Sets Syst. 100 (1–3) (1998) 327–342.
[55] D.S. Yeung, D.G. Chen, E.C.C. Tsang, J.W.T. Lee, X.Z. Wang, On the generalization of fuzzy rough sets, IEEE Trans. Fuzzy Syst. 13 (3) (2005) 343–361.
[56] M. Bernhard, On the T-transitivity of kernels, Fuzzy Sets Syst. 157 (2006) 787–1796.
[57] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognit. 40 (2007) 3509–3521.
[58] D.R. Wilson, T.R. Martinez, Improved heterogeneous distance functions, J. Artif. Intell. Res. 6 (1997) 1–34.