Research Article

A Novel Robust Fuzzy Rough Set Model for Feature Selection

Yuwen Li ¹, Shoushui Wei,² Xing Liu,³ and Zhimin Zhang ⁴

¹School of Instrument Science and Engineering, Southeast University, Nanjing 210000, China

²School of Control Science and Engineering, Shandong University, Jinan 250100, China

³Department of Anesthesiology, The Third Xiangya Hospital, Central South University, Changsha 410013, China

⁴Science and Technology on Information Systems Engineering Laboratory, The 28th Research Institute of CETC, Nanjing 210000, China

Correspondence should be addressed to Yuwen Li; liyuwen@seu.edu.cn and Zhimin Zhang; zmzsdu@163.com

Received 3 December 2020; Revised 23 December 2020; Accepted 30 December 2020; Published 31 January 2021

Academic Editor: Jianxin Li

Copyright © 2021 Yuwen Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The existing fuzzy rough set (FRS) models all believe that the decision attribute divides the sample set into several “clear” decision classes, and this data processing method makes the model sensitive to noise information when conducting feature selection. To solve this problem, this paper proposes a robust fuzzy rough set model (RS-FRS) based on representative samples. Firstly, the fuzzy membership degree of the samples is defined to reflect its fuzziness and uncertainty, and RS-FRS model is constructed to reduce the influence of the noise samples. RS-FRS model does not need to set parameters for the model in advance and can effectively reduce the complexity of the model and human intervention. On this basis, the related properties of RS-FRS model are studied, and the sample pair selection algorithm (SPS) based on RS-FRS is used for feature selection. In this paper, RS-FRS is tested and analysed on the open 12 datasets. The experimental results show that RS-FRS model proposed can effectively select the most relevant features and has certain robustness to the noise information. The proposed model has a good applicability for data processing and can effectively improve the performance of feature selection.

1. Introduction

In the current era of big data, the data scale is massive, and the presentation is high-dimensional. The high dimension of data representation is mainly due to that data often contains a large number of redundant or irrelevant features, resulting in excessively high data dimension, which seriously reduces the processing capacity and time efficiency of pattern classification as well as the resolution ability of decision making. High-dimensional data also makes the fast, timely, and accurate data mining task face great challenges. Therefore, how to effectively select features for these data has become one of the hot topics in the field of machine learning [1, 2]. The purpose of feature selection is to remove a large number of irrelevant and redundant features from the original feature set on the premise of ensuring the learning performance to find a set of feature subsets containing all or most of the classification information of the original feature space to reduce the impact of “dimension disaster” and

improve the learning performance. Therefore, feature selection (or attribute reduction) is very necessary, which has become a research hotspot in machine learning. Meanwhile, fuzzy rough set (FRS) theory is not only an objective and effective mathematical tool to deal with incomplete and uncertain information [3–7] but also a powerful and effective computing paradigm to realize feature selection [8–10]. In recent years, FRS theory has been widely concerned and applied in data mining, machine learning, pattern recognition, and other fields [11–13].

FRS can effectively deal with the fuzziness and vagueness of data. However, based on the upper and lower approximation of the classical FRS model, the nearest sample of the given target sample is used in the calculation, so the classical FRS model constrained by the nearest sample is extremely sensitive to noise.

In order to improve the robustness of classical FRS and reduce the influence of noise samples on the approximation under the model, many robust FRS models were proposed

[14–20]. An important application of FRS theory is feature selection, also known as attribute reduction, which maintains the inconsistency between feature and decision label. This inconsistency is manifested as two samples having the same feature value but different decision labels. Redundant or irrelevant features can be deleted to improve the classification prediction performance of learning algorithm and save running time and space, so that people can have a clearer understanding of the actual problems based on FRS. Feature selection based on FRS refers to the removal of redundant and irrelevant features from data without changing the ability of data classification. The advantage of FRS method is that in the process of data processing, only the information of the data itself can be utilized without any other prior knowledge and additional information. Generally speaking, there are two types of FRS-based feature selection methods: heuristic method based on dependency [21–23] and a structured method based on a discernibility matrix [24–28]. The dependency-based heuristic method uses the positive region and the dependency function [21–23] as the feature evaluation criterion and uses the heuristic search to obtain the feature subset. Specifically, FRS-based feature selection is calculated by using the forward or backward searching method on the premise of keeping a certain metric constant. Pawlak [1] introduced the concept of positive region when describing the consistency between features and decision label to select features. Hu and Cercone [23] designed an algorithm for calculation of reduction under the condition that the positive region of the decision was kept unchanged. Another kind of method is structured method based on discernible matrix. By introducing the identification matrix, a Boolean identification function is constructed, and the minimal form of the identification function is obtained by logical operation, thus all possible reduction results of the decision table are obtained. Based on the identification information in the identification matrix, many scholars have studied the computational problems of reduction [24–28]. For example, Yao and Zhao [24] transformed an identification matrix into its minimal form by designing an absorption matrix algorithm to calculate the reduction, so that the union of all elements constituted a reduction. Because this algorithm needs to calculate all the elements in the identifiable matrix, it will consume a lot of calculation time. In order to save a lot of running time and storage space, Chen et al. [27, 28] determined minimal elements of discernibility matrix through sample pair selection method (SPS) and designed a fast algorithm for calculation reduction based on minimal elements.

In our previous work [29], we proposed a novel FRS model, namely, fuzzy rough set, with representative sample (RS-FRS) model, which is constructed to reduce the influence of noisy samples. In the RS-FRS model, the fuzziness of sample membership is taken into consideration. Using fuzzy equivalent approximate space, other subsets of domain space can be approximated more precisely than those from a conventional FRS model. RS-FRS model does not require preset parameters. Our pilot study indicates that implementing such a new framework could reduce the complexity of the model and human intervention.

However, the previous work needs further research in several aspects. Firstly, the proposed theorem which supports the model is not thoroughly derived, which calls for additional theoretical derivation to strengthen the mathematical background. Moreover, the verification of the proposed model is not comprehensive since the model was tested using only the KNN classifier. Lastly, the previous work lacks solid statistical validations due to its nature as a pilot study that explores the method's potential effectiveness.

In this manuscript, we extensively addressed these drawback points. We conducted detailed derivations of Properties 2 and 3 to complete the RS-FRS model at the theoretical level. The completion of the theorem also supports the generation of the feature selection algorithm, which is significant to the framework based on RS-FRS models. In addition to the previous KNN classifier, the performance evaluation of the proposed method is extended by validating the algorithm using CART and LSVM classifiers. Our new results show satisfactory accuracy and robustness values, which further support this new method's effectiveness. Lastly, a comprehensive statistical analysis of the results is conducted and reported in this manuscript by applying Friedman tests and Bonferroni-Dunn tests to the model outputs. In summary, the previous pilot study is strengthened with theoretical, experimental, and statistical proofs to demonstrate the performance and robustness of the proposed RS-FRS method.

2. Preliminaries

The equivalence classes of Pawlak rough set are crisp subsets of the domain. These crisp information granules cannot reflect the fuzziness in reasoning. In practical classification learning, the features describing samples may be fuzzy, or the relations between samples are fuzzy relations calculated by numerical attributes. Therefore, FRS as Pawlak rough set extension model came into being.

For a nonempty universe U , if R is a binary relation and it satisfies reflexivity, symmetry, and sup-min transitivity, R is a fuzzy equivalence relation. The fuzzy equivalence class $[x]_R$ is generated by R with respect to sample $x \in U$. $[x]_R$ is a fuzzy set on U , which is also referred as the fuzzy neighborhood of x , i.e., $[x]_R(u) = R(x, u)$, for all $u \in U$.

Definition 1. Given a nonempty finite domain $U = \{x_1, x_2, \dots, x_n\}$, R is the fuzzy equivalence relation of U . For $\forall x \in U$, the fuzzy equivalence class of x is $[x]_R = r_{i1}/x_1 + r_{i2}/x_2 + \dots + r_{in}/x_n$, and it is the fuzzy subset of U . The membership degree value of $\forall x_j \in U$ to $[x]_R$ is $[x]_R(x_j) = r_{ij}$. The set of fuzzy equivalence classes forms a basic conceptual system for approximating any subset in the theoretic domain space, which is called fuzzy equivalence approximate space $FAS = \langle U, R \rangle$. The upper and lower approximations of $X \in F(U)$ in $\langle U, R \rangle$ are defined as

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

In equation (1), the upper approximation of the sample $x \in U$ is determined by $R(x, y)$ and $X(y)$. The lower approximation of the sample $x \in U$ is determined by $1 - R(x, y)$ and $X(y)$, where $X(y)$ is the membership degree of the sample $y \in U$ with respect to the equivalence class $X \in F(U)$.

At present, the existing fuzzy rough set models [14–20] consider that the decision attribute D divides the sample set into several crisp decision classes, so the membership degree $X(y)$ in the upper and lower approximations of the FRS model is a binary function with the value of 0 or 1. Then, the upper and lower approximations of the classical FRS model are reduced to

$$\begin{cases} \underline{R}D_i(x) = \inf_{u \notin D_i} \{1 - R(x, u)\}, \\ \overline{R}D_i(x) = \sup_{u \in D_i} \{R(x, u)\}. \end{cases} \quad (2)$$

3. The Proposed Model

At present, the existing FRS models [14–20] consider that the decision attribute D divides the sample set U into several crisp decision classes, so $X(y)$ as the membership degree in equation (1) is a binary function with the value of 0 or 1. Then, the upper and lower approximations of classical FRS model are degenerated as equation (2). However, this strategy that it divides the sample set into crisp decision classes is extremely sensitive to the noise samples, and if there exist noise samples, $X(y)$ has fuzzy uncertainty. Only defining $X(y)$ as a binary function of 0 or 1 does not meet the requirements of practical application, and it cannot well reflect the real membership relations between samples and each equivalence class. Therefore, determining the membership of samples is another important challenge for fuzzy rough set models.

3.1. Representative Sample. Unlike the existing FRS models, we consider D divides U into several fuzzy decision classes and define “representative sample” to calculate the fuzzy membership of samples. To be specific, the corresponding representative samples are found for each label. Then, the fuzzy membership degree of the target sample with respect to each label is calculated according to the distance between the target sample and the representative sample, so as to design a robust FRS model.

Definition 2 (see [29]). Let $\langle U, A \cup D \rangle$ be a fuzzy decision system, where the sample set $U = \{x_1, x_2, \dots, x_n\}$ has m attributes $A = \{a_1, a_2, \dots, a_m\}$. The decision attribute D divides the sample set U into r crisp equivalent decision classes $U/D = \{D_1, D_2, \dots, D_i, \dots, D_r\}$ ($1 \leq i \leq r$). The representative sample RS_i of the class D_i is defined as

$$RS_i = \arg_{x \in D_i} \min \sum_{y \in D_i} d(x, y), \quad (3)$$

where $d(x, y)$ is the distance between two samples in the class D_i . In this paper, Euclidean distance is used as a basic implementation of $d(x, y)$.

Definition 3 (see [29]). Let $\langle U, A \cup D \rangle$ be a fuzzy decision system, where the sample set $U = \{x_1, x_2, \dots, x_n\}$ has m attributes $A = \{a_1, a_2, \dots, a_m\}$. The decision attribute D divides the sample set U into r crisp equivalent decision classes $U/D = \{D_1, D_2, \dots, D_i, \dots, D_r\}$ ($1 \leq i \leq r$). The representative sample of the class D_i is RS_i . The membership degree of the sample $x \in U$ with respect to class D_i is defined as

$$D_i(x) = \frac{d(x, RS_i)}{\sum_{k=1}^r d(x, RS_k)}, \quad (4)$$

where $d(x, RS_i)$ is the distance between sample x and representative sample RS_i .

According to Definition 3, we can determine the membership degree of sample x with respect to each equivalence class by calculating the distance between sample x and the representative samples of each equivalence class. The membership degree $D_i(x)$ meets $D_i(x) \in (0, 1)$ and $\sum_{i=1}^r D_i(x) = 1$. It can be seen that $D_i(x)$ can fully reflect the fuzziness of sample membership. The larger the value of $D_i(x)$, the higher the degree of sample x belongs to class D_i . The smaller the value of $D_i(x)$, the lower the degree of sample x belongs to class D_i . In the dataset, the samples located in the boundary region may be the noise samples that have been mislabeled. By the calculation of fuzzy membership, the degree of the boundary samples belonging to each class can be determined.

3.2. FRS Model with Representative Sample. Based on Definitions 2 and 3, we propose a FRS model with representative samples (RS-FRS).

Definition 4 (see [29]). Let $\langle U, A \cup D \rangle$ be a fuzzy decision system, where the sample set $U = \{x_1, x_2, \dots, x_n\}$ has m attributes $A = \{a_1, a_2, \dots, a_m\}$. The decision attribute D divides the sample set U into r crisp equivalent decision classes $U/D = \{D_1, D_2, \dots, D_i, \dots, D_r\}$ ($1 \leq i \leq r$). R is a fuzzy equivalence relation on U . The representative sample of the class D_i is RS_i . The upper and lower approximations of RS-FRS model are defined as

$$\begin{aligned} \underline{R}^{RS} D_i(x) &= \inf_{u \in U} \max\{1 - R(x, u), D_i(u)\}, \\ \overline{R}^{RS} D_i(x) &= \sup_{u \in U} \min\{R(x, u), D_i(u)\}, \end{aligned} \quad (5)$$

where $D_i(u) = d(u, RS_i) / \sum_{k=1}^r d(u, RS_k)$.

The lower approximation of the FRS model indicates the certainty that a sample belongs to its decision class in the fuzzy approximation space, and the upper approximation indicates the possibility that a sample belongs to its decision class in the fuzzy approximation space. Therefore, the lower approximation can be used for classification and feature selection.

The samples of dataset located in the classification boundary region are most likely to be noise samples. Because

these samples are close to the samples of other classes, these noise samples are often used to calculate the lower approximation of the classical FRS model. This process causes the lower approximation to get smaller. In the proposed RS-FRS model, we consider not only the distance between the target sample and the nearest different classes' sample but also the fuzzy membership of the nearest different classes' sample. The fuzzy membership degree is calculated to expand the lower approximation of the model and reduce the influence of noise sample on the lower approximation, thus RS-FRS is robust.

The main difference between the RS-FRS model and the classical FRS is that the fuzziness of sample membership is ignored when the upper and lower approximations are calculated based on the classical FRS. This can easily lead to errors in the upper and lower approximations of classical FRS model when dealing with datasets containing noise samples. Therefore, the classical FRS can only maintain the maximum fuzzy dependence, but it cannot process noise information well. In contrast, the RS-FRS model considers the fuzziness of sample membership degree and can more precisely approximate other subsets of the domain space with fuzzy equivalent approximation space, so that the data fitting effect is better. Compared with the existing robust FRS model, the RS-FRS model does not need to set parameters for the model in advance, which can effectively reduce the model complexity and human intervention.

3.3. Related Properties of FRS Model with Representative Sample. For standard max operator $S(x, y) = \max(x, y)$, standard min operator $T(x, y) = \min(x, y)$, and standard complement operator $N(x) = 1 - x$, some properties of RS-FRS model are discussed. If other fuzzy operators are used [13], the relevant conclusions can be similarly generalized.

Property 1 (see[29]). For $\forall A \in F(U)$, the following statements hold:

$$\begin{aligned} (1) \quad \underline{R}^{\text{RS}} A &= N\left(\overline{R}^{\text{RS}}(A)\right), \\ (2) \quad \overline{R}^{\text{RS}} A &= N\left(\underline{R}^{\text{RS}}(A)\right). \end{aligned} \quad (6)$$

Proof. $\forall x \in U$

$$\begin{aligned} N\left(\overline{R}^{\text{RS}}(A(x))\right) &= N\left(\sup_{u \in U} \min\{R(x, u), A(u)\}\right) \\ &= \inf_{u \in U} (N(\min\{R(x, u), A(u)\})) \\ &= \inf_{u \in U} \max\{1 - R(x, u), A(u)\} \\ &= \underline{R}^{\text{RS}} A, \\ N\left(\underline{R}^{\text{RS}}(A(x))\right) &= N\left(\inf_{u \in U} \max\{1 - R(x, u), A(u)\}\right) \\ &= \sup_{u \in U} N(\max\{1 - R(x, u), A(u)\}) \\ &= \sup_{u \in U} \min\{R(x, u), A(u)\} \\ &= \overline{R}^{\text{RS}} A. \end{aligned} \quad (7)$$

□

Property 2. For $\forall A \in F(U)$, if x is a normal sample, the following statements hold:

$$\underline{R}^{\text{RS}} A \subseteq A \subseteq \overline{R}^{\text{RS}} A. \quad (8)$$

Proof. $\forall x \in U$

$$\begin{aligned} \underline{R}^{\text{RS}} A(x) &= \inf_{u \in U} \max\{1 - R(x, u), A(u)\} \\ &\leq \max\{1 - R(x, u), A(u)\}. \end{aligned} \quad (9)$$

Because x is a normal sample, x can be used directly to compute its lower approximation $\underline{R}^{\text{RS}} A(x)$,

$$\begin{aligned} \max\{1 - R(x, u), A(u)\} &= \max\{1 - R(x, x), A(x)\} \\ &= \max\{0, A(x)\} = A(x). \end{aligned} \quad (10)$$

Therefore, $\underline{R}^{\text{RS}} A \subseteq A$.

$$\begin{aligned} \overline{R}^{\text{RS}} A(x) &= \sup_{u \in U} \min\{R(x, u), A(u)\} \\ &\geq \min\{R(x, u), A(u)\}. \end{aligned} \quad (11)$$

Because x is a normal sample, x can be used directly to compute its upper approximation $\overline{R}^{\text{RS}} A(x)$,

$$\begin{aligned} \min\{R(x, u), D_i(u)\} &= \min\{R(x, x), A(x)\} \\ &= \min\{1, A(x)\} = A(x). \end{aligned} \quad (12)$$

Therefore, $A \subseteq \overline{R}^{\text{RS}} A$. □

Property 3. For $\forall A \in F(U)$, $\overline{R}_T A$ and $\underline{R}_S A$ are the upper and lower approximations of the classical FRS, respectively. The following statements hold:

$$\begin{aligned} (1) \quad \underline{R}^{\text{RS}} A &\supseteq \underline{R}_S A, \\ (2) \quad \overline{R}^{\text{RS}} A &\subseteq \overline{R}_T A. \end{aligned} \quad (13)$$

Proof. $\forall x \in U$, according to equation (2) in Section 2, when the standard operator is given, then

$$\underline{R}^{\text{RS}} A(x) = \inf_{u \in U} \max\{1 - R(x, u), A(u)\} \geq \inf\{1 - R(x, u)\} = \underline{R}_S A. \quad (14)$$

Therefore, $\underline{R}^{\text{RS}} A \supseteq \underline{R}_S A$.

$$\overline{R}^{\text{RS}} A(x) = \sup_{u \in U} \min\{R(x, u), A(u)\} \leq \sup\{R(x, u)\} = \overline{R}_T A. \quad (15)$$

Therefore, $\overline{R}^{\text{RS}} A \subseteq \overline{R}_T A$.

Property 1 proves the duality between the upper and lower approximations. Property 2 demonstrates the inclusion relationship between the upper and lower approximations and the original sample set. Property 3 proves the relationship of the upper and lower approximations between the RS-FRS model and the classical FRS model. The presentation and derivation of these properties provide a theoretical basis for the establishment and design of subsequent algorithms. □

4. Feature Selection Based on FRS Model with Representative Sample

Although the method based on discernibility matrix [27] can be used for feature selection, it has two obvious disadvantages: (1) a waste of computing time cost. All feature subsets can be obtained by using discernibility matrix method. However, when the result of feature selection is used for pattern recognition or classification, only one feature subset is sufficient. In other words, feature selection based on discernibility matrix method will increase invalid time cost. (2) A waste of storage space. When using the discernibility matrix, it is necessary to store all the elements in this matrix and reduce them through absorption operator. However, the key to find one feature subset is to determine the minimum elements. Compared with storing the minimum elements, the discernibility matrix method will cause a waste of storage space. Therefore, in this paper, we use the improved discernibility matrix method, named sample pair selection algorithm (SPS) [28]. Since each of these minimum elements is determined by at least one sample pair, SPS algorithm can select the pairs of samples corresponding to the minimum elements in the sample set. Based on this, we propose a feature selection algorithm based on the RS-FRS model with SPS to save time and space costs.

Let $\langle U, A \cup D \rangle$ be a fuzzy decision system, where the sample set $U = \{x_1, x_2, \dots, x_n\}$ has m attributes $A = \{a_1, a_2, \dots, a_m\}$. Each conditional attribute $a_t \in A$ induces a fuzzy relation $R_{\{a_t\}}$. Attribute subset $B \subseteq A$ induces the fuzzy relation $R_B = \bigcap_{a_t \in B} R_{\{a_t\}}$. The discernibility matrix of

this decision system is a $n \times n$ matrix, denoted by $M_D(U, A)$, and the element of discernibility matrix is c_{ij} :

$$\begin{aligned} (1) \quad c_{ij} &= \left\{ a_t : 1 - R_{\{a_t\}}(x_i, x_j) \geq \lambda_i \right\}, \\ \lambda_i &= \underline{R}_A[x_i]_D(x_i), \text{ if } x_j \notin [x_i]_D; \\ (2) \quad c_{ij} &= \emptyset, \quad \text{otherwise,} \end{aligned} \quad (16)$$

where λ_i is the lower approximation of x_i with respect to $[x_i]_D$.

In the RS-FRS model, the relative discrimination relation of conditional attribute a_t with respect to decision attribute D is defined as a binary relation, which is calculated by the following equation:

$$\text{DIS}'(R_{\{a_t\}}) = \left\{ (x_i, x_j) \in U \times U \mid 1 - R_{\{a_t\}}(x_i, x_j) \geq \lambda_i, x_i \notin [x_j]_D \right\}, \quad (17)$$

where $\lambda_i = \underline{R}_A^{SR}[x_i]_D(x_i)$ and $\text{DIS}'(R_A) = \bigcup_{a_t \in A} \text{DIS}'(R_{\{a_t\}})$.

We obtain the following results by finding the relationship between $\text{DIS}'(R_{\{a_t\}})$ and element c_{ij} in the discernible matrix $M_D(U, A)$. If $c_{ij} \neq \emptyset$, $(x_i, x_j) \in \text{DIS}'(R_{\{a_t\}}) \iff a_t \in c_{ij}$. $N_{ij} = |\left\{ a_t \mid (tx_i, n, qx_j, h) \in x\text{DIS}'IC(; R_{\{a_t\}}) \right\}|$, where N_{ij} is the number of conditional attributes that satisfied $(x_i, x_j) \in \text{DIS}'(R_{\{a_t\}})$. Obviously, $N_{ij} = |c_{ij}|$.

According to the above definitions and analysis, a feature selection algorithm based on RS-FRS model with SPS can be described by Algorithm 1.

5. Experiments

In this section, we compare the proposed RS-FRS model with several existing models from the three aspects of classification accuracy, robustness, and statistical properties. The existing models include FRS [30], β -PFRS ($\beta = 0.5$) [14], K-trimmed FRS ($k = 5$) [18], K-means FRS ($k = 5$) [18], K-median FRS ($k = 5$) [18], and SFRS ($r = 0.1$) [15]. To test the performance of all FRS models in feature selection, these experiments are conducted on three common classifiers, including k -nearest neighbor (KNN, $K = 3$) [31], classification regression tree (CART) [32], and linear support vector machine (LSVM) [33].

5.1. Datasets. In our experiments, 12 datasets from the open access UCI database [34] are utilized and described in Table 1. UCI dataset is a commonly used dataset of machine learning standard tests. For example, in dataset "wine", these data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. The attributes are alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue, OD280/OD315 of diluted wines, and proline.

5.2. Results

5.2.1. Comparison of Classification Accuracy. On three classifiers, we conduct comparison of classification accuracy with different FRS models in original data and noise data. The corresponding noise levels are 0%, 5%, and 10%, respectively. 0% noise level means the original data (the original data itself assumes no noise).

5% (or 10%) noise level means there are 5% (or 10%) samples mislabeled randomly in the original data. In order to ensure the reliability of the experiments, we carry out the 10-fold cross-validation based on each independent noise random process. Tables 2–4 show the comparison of classification performance of different FRS models on KNN, CART, and LSVM in original and noisy data. Bold values indicate the optimal performance among all FRS models, and italic values indicate the average classification accuracy of each model under all datasets.

In terms of the global perspective on the three classifiers, the average precision of the RS-FRS model is higher than that of the other FRS models upon the original data (0%) and the noise data (5% and 10%). Furthermore, in terms of the perspective of each dataset, each classifier, and each noise level, with KNN, RS-FRS model is optimal on 8 out of 12 datasets upon the original data (0%), and it is optimal on 15 out of 24 datasets upon the noise data (5% and 10%) in Table 2. Similar results are shown in Tables 3 and 4. With CART, RS-FRS model is superior on 7 out of 12 datasets upon the original data (0%), and it also is superior on 16 out of 24 datasets upon the noise data (5% and 10%) in Table 3. With LSVM, RS-FRS model is optimal on 6 out of 12

Input: a set of condition attributes A ; a set of samples U .

Output: selected feature subset S

- (1) $\forall x_i \in U$, according to equation (5), the lower approximation λ_i of RS-FRS model is computed;
- (2) Compute every $\text{DIS}'(R_{\{a_i\}})$ and $\text{DIS}'(R_A)$;
- (3) Sort $(x_i, x_j) \in \text{DIS}'(R_{\{a_i\}})$ according to N_{ij} ;
- (4) **while** $\text{DIS}'(R_A) \neq \emptyset$ **do**
- (5) Selected the first sample pair $(x_{i0}, x_{j0}) \in \text{DIS}'(R_A)$;
- (6) Select one a_t such that $(x_i, x_j) \in \text{DIS}'(R_{\{a_t\}})$ and add a_t into S ;
- (7) $\text{DIS}'(R_A) = \text{DIS}'(R_A) - \text{DIS}'(R_{\{a_t\}})$;
- (8) **end while**
- (9) **return** S

ALGORITHM 1: Feature selection algorithm based on RS-FRS model with SPS.

TABLE 1: Descriptions of 12 datasets.

No.	Datasets	Samples	Attributes	Classes
1	Glass	214	9	6
2	Wine	178	13	3
3	Heart	270	13	2
4	Segment	2310	18	7
5	Hepatitis	155	19	2
6	ICU	200	20	3
7	German	1000	20	2
8	Soy	47	21	4
9	Horse	368	22	2
10	WDBC	569	30	2
11	WPBC	198	33	2
12	Sonar	208	60	2

datasets upon the original data (0%), and it is optimal on 16 out of 24 datasets upon the noise data (5% and 10%) in Table 4. It is worth noting that for wine, soy, hepatitis, ICU and WPBC, these 7 FRS models have the same or similar performance in 0% noise level. This is because there is no noise in the original data according to our obvious assumption; the 7 FRS models find the same sample as the nearest sample of the target sample to calculate the upper and lower approximations of the target sample.

5.2.2. Robustness Analysis. In addition, after adding the noise gradually, the decreasing degree of the average precision of different models (Tables 2–4) is used as the basis for comparing the robustness of these models. In Figures 1–3, FRS, β -PFRS, K -trimmed FRS, K -means FRS, K -median FRS, SFRS, and RS-FRS model are numbered from 1 to 7 in sequence.

As shown in Figures 1–3, as the noise level increases gradually, the performance of every FRS model declines on three classifiers. This kind of phenomenon is consistent with the actual situation, because the noisy sample may be regarded as the nearest sample of the given target sample when these models calculate the lower approximation of the target sample. Therefore, the classification accuracy of these FRS models will reduce. However, among these models, the classical FRS model declines more sharply and changes more dramatically. This is because the classical FRS model is the most dependent on the nearest sample and it has the least

robust performance. The proposed RS-FRS model presents the smallest decline and the least drastic change, indicating that our strategy that divides the sample set into fuzzy decision classes is in line with reality and has the best robust performance.

5.2.3. Statistical Analysis. To further explore whether there are significant differences in the average classification performance of the seven FRS models, we adopted Friedman test [35] and Bonferroni-Dunn test [36]. The Friedman statistical coefficient is defined as

$$\chi_F^2 = \frac{12N}{k(k+1)} \left(\sum_{i=1}^k R_i^2 - \frac{k(k+1)^2}{4} \right), \quad (18)$$

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

where R_i is the average ranking of model i in all datasets and k is the total number of models and N is the number of datasets. F_F obeys the Fisher distribution of $(k-1)$ and $(N-1)(N-1)$ degrees of freedom. In this paper, $k=7$ and $N=36$. At significance level $\alpha=0.1$, $F(6, 210)=1.77$. Table 5 shows the Friedman statistical coefficients and corresponding critical values under different classifiers. If F_F is greater than the critical value, the null hypothesis is rejected, and we believe that there is a significant difference among the performance of all models. On the contrary, the null hypothesis cannot be rejected and it can be considered that there is no significant difference among the performances of all models.

If Friedman's null hypothesis is rejected, the Bonferroni-Dunn statistical verification method can further analyze the relative performance between each comparison model and RS-FRS model, and RS-FRS model is regarded as the control model. If there is a significant difference between the two models, the difference between the average ranking between them should be at least greater than the critical difference (CD):

$$\text{CD} = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \quad (19)$$

TABLE 2: Comparison of classification performance of different FRS models on KNN in original and noisy data.

Datasets	Noise level (%)	FRS	β -PFRS	K -trimmed FRS	K -means FRS	K -median FRS	SFRS	RS-FRS
Glass	0	0.678 ± 0.134	0.650 ± 0.068	0.631 ± 0.061	0.650 ± 0.078	0.655 ± 0.099	0.678 ± 0.134	0.668 ± 0.031
	5	0.629 ± 0.061	0.649 ± 0.027	0.630 ± 0.037	0.647 ± 0.091	0.648 ± 0.043	0.655 ± 0.096	0.656 ± 0.029
	10	0.628 ± 0.098	0.647 ± 0.066	0.626 ± 0.110	0.645 ± 0.019	0.644 ± 0.068	0.648 ± 0.089	0.649 ± 0.015
Wine	0	0.931 ± 0.072	0.926 ± 0.054	0.926 ± 0.054	0.931 ± 0.072	0.931 ± 0.072	0.931 ± 0.072	0.931 ± 0.072
	5	0.892 ± 0.069	0.910 ± 0.045	0.925 ± 0.059	0.930 ± 0.059	0.924 ± 0.062	0.920 ± 0.073	0.928 ± 0.086
	10	0.876 ± 0.058	0.909 ± 0.038	0.924 ± 0.035	0.912 ± 0.072	0.908 ± 0.054	0.919 ± 0.071	0.926 ± 0.044
Heart	0	0.744 ± 0.069	0.759 ± 0.082	0.741 ± 0.105	0.730 ± 0.078	0.730 ± 0.078	0.744 ± 0.069	0.744 ± 0.069
	5	0.735 ± 0.071	0.739 ± 0.059	0.740 ± 0.110	0.728 ± 0.080	0.727 ± 0.013	0.741 ± 0.067	0.742 ± 0.038
	10	0.719 ± 0.082	0.733 ± 0.086	0.734 ± 0.044	0.726 ± 0.019	0.727 ± 0.078	0.736 ± 0.054	0.736 ± 0.025
Segment	0	0.946 ± 0.109	0.949 ± 0.109	0.953 ± 0.109	0.955 ± 0.109	0.953 ± 0.109	0.946 ± 0.033	0.946 ± 0.109
	5	0.944 ± 0.038	0.948 ± 0.036	0.949 ± 0.032	0.951 ± 0.092	0.952 ± 0.034	0.945 ± 0.055	0.945 ± 0.010
	10	0.941 ± 0.040	0.947 ± 0.032	0.946 ± 0.053	0.949 ± 0.039	0.947 ± 0.037	0.944 ± 0.071	0.944 ± 0.104
Hepatitis	0	0.765 ± 0.107	0.765 ± 0.107	0.765 ± 0.107	0.765 ± 0.107	0.765 ± 0.107	0.765 ± 0.107	0.765 ± 0.107
	5	0.763 ± 0.092	0.764 ± 0.108	0.763 ± 0.045	0.764 ± 0.099	0.760 ± 0.123	0.764 ± 0.092	0.764 ± 0.062
	10	0.750 ± 0.118	0.759 ± 0.075	0.760 ± 0.075	0.764 ± 0.081	0.753 ± 0.077	0.750 ± 0.028	0.762 ± 0.104
ICU	0	0.868 ± 0.162	0.879 ± 0.183	0.879 ± 0.183	0.868 ± 0.129	0.868 ± 0.129	0.868 ± 0.162	0.880 ± 0.042
	5	0.856 ± 0.176	0.859 ± 0.087	0.860 ± 0.158	0.857 ± 0.162	0.855 ± 0.146	0.861 ± 0.054	0.863 ± 0.107
	10	0.850 ± 0.023	0.855 ± 0.126	0.853 ± 0.101	0.856 ± 0.023	0.852 ± 0.176	0.856 ± 0.010	0.862 ± 0.093
German	0	0.716 ± 0.037	0.712 ± 0.061	0.713 ± 0.045	0.713 ± 0.032	0.713 ± 0.032	0.716 ± 0.037	0.716 ± 0.037
	5	0.708 ± 0.038	0.710 ± 0.046	0.710 ± 0.059	0.709 ± 0.037	0.712 ± 0.041	0.713 ± 0.063	0.714 ± 0.031
	10	0.706 ± 0.033	0.710 ± 0.035	0.708 ± 0.042	0.708 ± 0.091	0.709 ± 0.032	0.710 ± 0.025	0.713 ± 0.088
Soy	0	0.850 ± 0.141	0.825 ± 0.134	0.850 ± 0.141	0.850 ± 0.141	0.850 ± 0.141	0.850 ± 0.141	0.850 ± 0.141
	5	0.790 ± 0.062	0.819 ± 0.077	0.818 ± 0.025	0.818 ± 0.018	0.816 ± 0.050	0.820 ± 0.041	0.820 ± 0.112
	10	0.760 ± 0.046	0.818 ± 0.082	0.810 ± 0.082	0.806 ± 0.087	0.809 ± 0.246	0.816 ± 0.092	0.818 ± 0.070
Horse	0	0.864 ± 0.056	0.845 ± 0.051	0.834 ± 0.060	0.834 ± 0.060	0.840 ± 0.061	0.845 ± 0.051	0.864 ± 0.056
	5	0.830 ± 0.042	0.829 ± 0.061	0.823 ± 0.072	0.830 ± 0.055	0.825 ± 0.073	0.828 ± 0.052	0.828 ± 0.081
	10	0.815 ± 0.056	0.825 ± 0.048	0.818 ± 0.065	0.829 ± 0.061	0.824 ± 0.074	0.826 ± 0.014	0.827 ± 0.099
WDBC	0	0.937 ± 0.036	0.955 ± 0.019	0.956 ± 0.023	0.946 ± 0.034	0.952 ± 0.034	0.937 ± 0.028	0.937 ± 0.036
	5	0.928 ± 0.032	0.933 ± 0.041	0.951 ± 0.020	0.941 ± 0.037	0.952 ± 0.029	0.936 ± 0.071	0.936 ± 0.111
	10	0.920 ± 0.027	0.927 ± 0.021	0.949 ± 0.038	0.940 ± 0.034	0.951 ± 0.067	0.935 ± 0.083	0.935 ± 0.027
WPBC	0	0.757 ± 0.050	0.727 ± 0.083	0.721 ± 0.101	0.711 ± 0.095	0.721 ± 0.082	0.726 ± 0.101	0.757 ± 0.050
	5	0.721 ± 0.087	0.721 ± 0.066	0.712 ± 0.114	0.711 ± 0.098	0.711 ± 0.069	0.722 ± 0.081	0.751 ± 0.114
	10	0.720 ± 0.089	0.720 ± 0.053	0.710 ± 0.095	0.709 ± 0.108	0.706 ± 0.101	0.719 ± 0.077	0.747 ± 0.093
Sonar	0	0.731 ± 0.067	0.774 ± 0.060	0.779 ± 0.083	0.769 ± 0.098	0.769 ± 0.098	0.731 ± 0.067	0.780 ± 0.091
	5	0.730 ± 0.063	0.750 ± 0.055	0.763 ± 0.064	0.765 ± 0.106	0.764 ± 0.077	0.729 ± 0.013	0.772 ± 0.123
	10	0.728 ± 0.067	0.744 ± 0.089	0.726 ± 0.070	0.726 ± 0.075	0.730 ± 0.059	0.728 ± 0.088	0.761 ± 0.105
Average		0.798 ± 0.072	0.805 ± 0.069	0.804 ± 0.074	0.804 ± 0.074	0.804 ± 0.080	0.804 ± 0.068	0.812 ± 0.073

where k is the total number of models, N is the number of datasets, and q_α is the critical value of Bonferroni-Dunn test at the corresponding significance level. At significance level $\alpha = 0.1$, $q_\alpha = 2.394$. Therefore, $CD = 1.2190$ ($k = 7$, $N = 36$). If the average ranking difference between the RS-FRS model and a comparison model is greater than the value of CD, we consider their performance to be significantly different.

Figures 4–6 show the CD diagram of the three classifiers, in which the average ranking of all algorithms is drawn in turn along the horizontal axis. In other words, scales from 1 to 7 on the horizontal axis are the average ranking of the seven algorithms over 12 datasets, with the higher the ranking, the better the algorithm, and the algorithm with the highest ranking (optimal) in the axis is located on the right side of the axis. If the average ranking between RS-FRS and a comparison algorithm is connected by a thick line (CD value), it means that RS-FRS has comparable performance with the comparison algorithm. If the average ranking

difference between the RS-FRS model and a comparison model is greater than CD value, we consider their performance to be significantly different. As shown in the above figures, we can obtain the following results: (1) under the KNN, the RS-FRS is obviously better than the model of FRS, β -PFRS, K -trimmed FRS, K -means FRS, K -median FRS, and SFRS model. Moreover, no other comparison model has the same performance as RS-FRS model in statistical verification. (2) Under the CART, the RS-FRS is obviously better than the model of FRS, β -PFRS, K -trimmed FRS, K -means FRS, K -median FRS, and SFRS model. No other comparison model has the same performance as RS-FRS model in statistical verification. (3) Under the LSVM, the RS-FRS and SFRS models are solid strong to the FRS, β -PFRS, K -trimmed FRS, K -means FRS, and K -median FRS model. Moreover, RS-FRS and SFRS models have considerable performance on statistics. Through the analysis of the above, the RS-FRS model performance is better than that of FRS,

TABLE 3: Comparison of classification performance of different FRS models on CART in original and noisy data.

Datasets	Noise level (%)	FRS	β -PFRS	K -trimmed FRS	K -means FRS	K -median FRS	SFRS	RS-FRS
Glass	0	0.478 $\pm$ 0.031	0.436 $\pm$ 0.027	0.454 $\pm$ 0.039	0.458 $\pm$ 0.045	0.455 $\pm$ 0.025	0.478 $\pm$ 0.031	0.478 $\pm$ 0.031
	5	0.430 $\pm$ 0.025	0.434 $\pm$ 0.026	0.433 $\pm$ 0.010	0.436 $\pm$ 0.004	0.434 $\pm$ 0.049	0.463 $\pm$ 0.069	0.465 $\pm$ 0.060
	10	0.429 $\pm$ 0.069	0.431 $\pm$ 0.014	0.432 $\pm$ 0.025	0.431 $\pm$ 0.051	0.430 $\pm$ 0.052	0.454 $\pm$ 0.010	0.460 $\pm$ 0.025
Wine	0	0.909 $\pm$ 0.068	0.909 $\pm$ 0.068	0.908 $\pm$ 0.081	0.908 $\pm$ 0.081	0.908 $\pm$ 0.081	0.908 $\pm$ 0.081	0.909 $\pm$ 0.068
	5	0.859 $\pm$ 0.059	0.903 $\pm$ 0.092	0.901 $\pm$ 0.026	0.860 $\pm$ 0.075	0.899 $\pm$ 0.064	0.902 $\pm$ 0.072	0.903 $\pm$ 0.083
	10	0.819 $\pm$ 0.076	0.857 $\pm$ 0.037	0.897 $\pm$ 0.087	0.859 $\pm$ 0.041	0.893 $\pm$ 0.056	0.896 $\pm$ 0.054	0.902 $\pm$ 0.107
Heart	0	0.707 $\pm$ 0.019	0.759 $\pm$ 0.099	0.752 $\pm$ 0.089	0.707 $\pm$ 0.019	0.707 $\pm$ 0.019	0.707 $\pm$ 0.124	0.752 $\pm$ 0.052
	5	0.694 $\pm$ 0.054	0.737 $\pm$ 0.091	0.707 $\pm$ 0.019	0.702 $\pm$ 0.047	0.704 $\pm$ 0.010	0.705 $\pm$ 0.055	0.750 $\pm$ 0.049
	10	0.659 $\pm$ 0.045	0.728 $\pm$ 0.043	0.693 $\pm$ 0.011	0.682 $\pm$ 0.044	0.693 $\pm$ 0.031	0.704 $\pm$ 0.023	0.748 $\pm$ 0.076
Segment	0	0.947 $\pm$ 0.029	0.946 $\pm$ 0.039	0.945 $\pm$ 0.035	0.945 $\pm$ 0.035	0.945 $\pm$ 0.035	0.947 $\pm$ 0.029	0.947 $\pm$ 0.029
	5	0.941 $\pm$ 0.031	0.953 $\pm$ 0.027	0.942 $\pm$ 0.041	0.943 $\pm$ 0.034	0.942 $\pm$ 0.027	0.945 $\pm$ 0.083	0.946 $\pm$ 0.082
	10	0.938 $\pm$ 0.036	0.938 $\pm$ 0.030	0.940 $\pm$ 0.032	0.941 $\pm$ 0.035	0.941 $\pm$ 0.038	0.942 $\pm$ 0.064	0.944 $\pm$ 0.033
Hepatitis	0	0.800 $\pm$ 0.063	0.800 $\pm$ 0.063	0.800 $\pm$ 0.063	0.800 $\pm$ 0.063	0.800 $\pm$ 0.063	0.800 $\pm$ 0.063	0.800 $\pm$ 0.063
	5	0.762 $\pm$ 0.093	0.761 $\pm$ 0.103	0.770 $\pm$ 0.130	0.753 $\pm$ 0.081	0.755 $\pm$ 0.101	0.796 $\pm$ 0.039	0.798 $\pm$ 0.080
	10	0.723 $\pm$ 0.126	0.752 $\pm$ 0.079	0.732 $\pm$ 0.021	0.751 $\pm$ 0.104	0.728 $\pm$ 0.120	0.781 $\pm$ 0.041	0.790 $\pm$ 0.105
ICU	0	0.799 $\pm$ 0.085	0.783 $\pm$ 0.036	0.783 $\pm$ 0.036	0.783 $\pm$ 0.036	0.783 $\pm$ 0.036	0.799 $\pm$ 0.085	0.799 $\pm$ 0.085
	5	0.760 $\pm$ 0.062	0.764 $\pm$ 0.024	0.763 $\pm$ 0.022	0.773 $\pm$ 0.045	0.761 $\pm$ 0.019	0.762 $\pm$ 0.077	0.786 $\pm$ 0.091
	10	0.751 $\pm$ 0.013	0.755 $\pm$ 0.088	0.754 $\pm$ 0.096	0.755 $\pm$ 0.019	0.754 $\pm$ 0.049	0.752 $\pm$ 0.082	0.761 $\pm$ 0.114
German	0	0.716 $\pm$ 0.037	0.704 $\pm$ 0.035	0.713 $\pm$ 0.045	0.713 $\pm$ 0.032	0.713 $\pm$ 0.032	0.712 $\pm$ 0.046	0.716 $\pm$ 0.037
	5	0.716 $\pm$ 0.053	0.700 $\pm$ 0.031	0.711 $\pm$ 0.039	0.712 $\pm$ 0.034	0.710 $\pm$ 0.038	0.710 $\pm$ 0.071	0.713 $\pm$ 0.064
	10	0.707 $\pm$ 0.033	0.700 $\pm$ 0.061	0.709 $\pm$ 0.031	0.708 $\pm$ 0.042	0.709 $\pm$ 0.032	0.710 $\pm$ 0.079	0.709 $\pm$ 0.088
Soy	0	0.915 $\pm$ 0.111	0.915 $\pm$ 0.111	0.915 $\pm$ 0.111	0.915 $\pm$ 0.111	0.915 $\pm$ 0.111	0.915 $\pm$ 0.111	0.915 $\pm$ 0.111
	5	0.810 $\pm$ 0.067	0.883 $\pm$ 0.077	0.879 $\pm$ 0.004	0.891 $\pm$ 0.067	0.878 $\pm$ 0.058	0.894 $\pm$ 0.068	0.893 $\pm$ 0.108
	10	0.790 $\pm$ 0.039	0.810 $\pm$ 0.051	0.806 $\pm$ 0.058	0.808 $\pm$ 0.013	0.806 $\pm$ 0.025	0.821 $\pm$ 0.022	0.825 $\pm$ 0.024
Horse	0	0.855 $\pm$ 0.046	0.862 $\pm$ 0.049	0.857 $\pm$ 0.050	0.854 $\pm$ 0.047	0.857 $\pm$ 0.050	0.856 $\pm$ 0.052	0.855 $\pm$ 0.046
	5	0.846 $\pm$ 0.067	0.860 $\pm$ 0.053	0.854 $\pm$ 0.060	0.853 $\pm$ 0.050	0.852 $\pm$ 0.047	0.855 $\pm$ 0.013	0.855 $\pm$ 0.103
	10	0.841 $\pm$ 0.047	0.848 $\pm$ 0.055	0.848 $\pm$ 0.038	0.851 $\pm$ 0.033	0.853 $\pm$ 0.040	0.853 $\pm$ 0.051	0.854 $\pm$ 0.027
WDBC	0	0.907 $\pm$ 0.037	0.931 $\pm$ 0.046	0.907 $\pm$ 0.042	0.926 $\pm$ 0.044	0.916 $\pm$ 0.038	0.907 $\pm$ 0.037	0.928 $\pm$ 0.029
	5	0.903 $\pm$ 0.084	0.924 $\pm$ 0.031	0.903 $\pm$ 0.041	0.909 $\pm$ 0.033	0.914 $\pm$ 0.034	0.905 $\pm$ 0.102	0.927 $\pm$ 0.057
	10	0.895 $\pm$ 0.058	0.923 $\pm$ 0.022	0.905 $\pm$ 0.045	0.907 $\pm$ 0.032	0.896 $\pm$ 0.036	0.904 $\pm$ 0.046	0.924 $\pm$ 0.037
WPBC	0	0.660 $\pm$ 0.115	0.703 $\pm$ 0.128	0.645 $\pm$ 0.109	0.660 $\pm$ 0.134	0.634 $\pm$ 0.159	0.666 $\pm$ 0.090	0.660 $\pm$ 0.092
	5	0.630 $\pm$ 0.118	0.702 $\pm$ 0.112	0.640 $\pm$ 0.105	0.656 $\pm$ 0.153	0.631 $\pm$ 0.109	0.666 $\pm$ 0.086	0.659 $\pm$ 0.067
	10	0.620 $\pm$ 0.110	0.659 $\pm$ 0.104	0.629 $\pm$ 0.150	0.640 $\pm$ 0.135	0.629 $\pm$ 0.139	0.664 $\pm$ 0.053	0.659 $\pm$ 0.035
Sonar	0	0.668 $\pm$ 0.065	0.686 $\pm$ 0.122	0.688 $\pm$ 0.115	0.659 $\pm$ 0.066	0.659 $\pm$ 0.066	0.686 $\pm$ 0.065	0.668 $\pm$ 0.081
	5	0.645 $\pm$ 0.035	0.678 $\pm$ 0.099	0.654 $\pm$ 0.115	0.644 $\pm$ 0.035	0.643 $\pm$ 0.047	0.659 $\pm$ 0.012	0.666 $\pm$ 0.109
	10	0.634 $\pm$ 0.105	0.659 $\pm$ 0.106	0.647 $\pm$ 0.110	0.638 $\pm$ 0.088	0.636 $\pm$ 0.095	0.658 $\pm$ 0.025	0.665 $\pm$ 0.053
Average		0.755 $\pm$ 0.061	0.772 $\pm$ 0.063	0.764 $\pm$ 0.060	0.762 $\pm$ 0.056	0.761 $\pm$ 0.056	0.772 $\pm$ 0.059	0.779 $\pm$ 0.067

TABLE 4: Comparison of classification performance of different FRS models on LSVM in original and noisy data.

Datasets	Noise level (%)	FRS	β -PFRS	K -trimmed FRS	K -means FRS	K -median FRS	SFRS	RS-FRS
Glass	0	0.548 $\pm$ 0.122	0.571 $\pm$ 0.126	0.570 $\pm$ 0.128	0.581 $\pm$ 0.128	0.575 $\pm$ 0.128	0.582 $\pm$ 0.122	0.548 $\pm$ 0.122
	5	0.542 $\pm$ 0.034	0.569 $\pm$ 0.021	0.567 $\pm$ 0.007	0.575 $\pm$ 0.033	0.574 $\pm$ 0.051	0.580 $\pm$ 0.085	0.547 $\pm$ 0.090
	10	0.539 $\pm$ 0.028	0.564 $\pm$ 0.015	0.565 $\pm$ 0.027	0.571 $\pm$ 0.019	0.572 $\pm$ 0.029	0.577 $\pm$ 0.060	0.545 $\pm$ 0.110
Wine	0	0.938 $\pm$ 0.055	0.950 $\pm$ 0.072	0.950 $\pm$ 0.072	0.938 $\pm$ 0.055	0.938 $\pm$ 0.055	0.948 $\pm$ 0.055	0.949 $\pm$ 0.068
	5	0.892 $\pm$ 0.062	0.945 $\pm$ 0.055	0.936 $\pm$ 0.079	0.937 $\pm$ 0.039	0.935 $\pm$ 0.050	0.946 $\pm$ 0.034	0.947 $\pm$ 0.101
	10	0.890 $\pm$ 0.064	0.940 $\pm$ 0.027	0.935 $\pm$ 0.070	0.933 $\pm$ 0.029	0.935 $\pm$ 0.056	0.941 $\pm$ 0.082	0.942 $\pm$ 0.069
Heart	0	0.770 $\pm$ 0.087	0.830 $\pm$ 0.058	0.778 $\pm$ 0.070	0.767 $\pm$ 0.080	0.767 $\pm$ 0.080	0.770 $\pm$ 0.087	0.770 $\pm$ 0.087
	5	0.767 $\pm$ 0.080	0.781 $\pm$ 0.044	0.772 $\pm$ 0.086	0.761 $\pm$ 0.051	0.760 $\pm$ 0.039	0.768 $\pm$ 0.066	0.766 $\pm$ 0.043
	10	0.753 $\pm$ 0.043	0.754 $\pm$ 0.055	0.752 $\pm$ 0.065	0.741 $\pm$ 0.012	0.752 $\pm$ 0.025	0.755 $\pm$ 0.038	0.765 $\pm$ 0.071
Segment	0	0.903 $\pm$ 0.055	0.904 $\pm$ 0.054	0.904 $\pm$ 0.053	0.905 $\pm$ 0.052	0.904 $\pm$ 0.053	0.905 $\pm$ 0.052	0.905 $\pm$ 0.052
	5	0.894 $\pm$ 0.057	0.896 $\pm$ 0.018	0.901 $\pm$ 0.087	0.895 $\pm$ 0.010	0.895 $\pm$ 0.071	0.899 $\pm$ 0.025	0.903 $\pm$ 0.044
	10	0.874 $\pm$ 0.020	0.876 $\pm$ 0.042	0.875 $\pm$ 0.059	0.877 $\pm$ 0.076	0.878 $\pm$ 0.093	0.894 $\pm$ 0.101	0.897 $\pm$ 0.069
Hepatitis	0	0.815 $\pm$ 0.069	0.815 $\pm$ 0.080	0.815 $\pm$ 0.080	0.815 $\pm$ 0.080	0.815 $\pm$ 0.080	0.845 $\pm$ 0.080	0.845 $\pm$ 0.080
	5	0.805 $\pm$ 0.057	0.811 $\pm$ 0.036	0.809 $\pm$ 0.061	0.814 $\pm$ 0.027	0.813 $\pm$ 0.054	0.832 $\pm$ 0.099	0.834 $\pm$ 0.096
	10	0.793 $\pm$ 0.091	0.801 $\pm$ 0.061	0.815 $\pm$ 0.045	0.812 $\pm$ 0.052	0.811 $\pm$ 0.059	0.833 $\pm$ 0.067	0.833 $\pm$ 0.013

TABLE 4: Continued.

Datasets	Noise level (%)	FRS	β -PFRS	K -trimmed FRS	K -means FRS	K -median FRS	SFRS	RS-FRS
ICU	0	0.926 ± 0.023	0.926 ± 0.023	0.926 ± 0.023	0.926 ± 0.023	0.926 ± 0.023	0.926 ± 0.023	0.926 ± 0.023
	5	0.910 ± 0.047	0.920 ± 0.018	0.920 ± 0.023	0.923 ± 0.033	0.919 ± 0.039	0.921 ± 0.071	0.924 ± 0.104
	10	0.899 ± 0.013	0.911 ± 0.047	0.910 ± 0.043	0.903 ± 0.022	0.906 ± 0.019	0.919 ± 0.045	0.922 ± 0.055
German	0	0.735 ± 0.054	0.739 ± 0.057	0.739 ± 0.057	0.742 ± 0.055	0.742 ± 0.055	0.732 ± 0.054	0.735 ± 0.054
	5	0.720 ± 0.055	0.722 ± 0.104	0.725 ± 0.046	0.725 ± 0.115	0.729 ± 0.048	0.727 ± 0.015	0.731 ± 0.080
	10	0.701 ± 0.047	0.714 ± 0.054	0.711 ± 0.048	0.702 ± 0.044	0.708 ± 0.041	0.719 ± 0.039	0.725 ± 0.011
Soy	0	0.765 ± 0.024	0.775 ± 0.024	0.765 ± 0.024	0.765 ± 0.024	0.765 ± 0.024	0.765 ± 0.063	0.775 ± 0.024
	5	0.705 ± 0.032	0.740 ± 0.025	0.736 ± 0.039	0.740 ± 0.033	0.738 ± 0.074	0.739 ± 0.017	0.763 ± 0.117
	10	0.700 ± 0.047	0.731 ± 0.063	0.724 ± 0.032	0.722 ± 0.018	0.721 ± 0.058	0.735 ± 0.023	0.753 ± 0.027
Horse	0	0.832 ± 0.052	0.829 ± 0.052	0.842 ± 0.057	0.842 ± 0.057	0.834 ± 0.053	0.839 ± 0.052	0.832 ± 0.052
	5	0.826 ± 0.059	0.828 ± 0.069	0.834 ± 0.053	0.835 ± 0.044	0.831 ± 0.034	0.834 ± 0.073	0.831 ± 0.054
	10	0.807 ± 0.044	0.824 ± 0.068	0.832 ± 0.057	0.823 ± 0.050	0.821 ± 0.059	0.829 ± 0.026	0.830 ± 0.080
WDBC	0	0.935 ± 0.035	0.943 ± 0.020	0.935 ± 0.046	0.933 ± 0.047	0.939 ± 0.047	0.935 ± 0.035	0.935 ± 0.035
	5	0.930 ± 0.042	0.942 ± 0.025	0.933 ± 0.032	0.931 ± 0.035	0.933 ± 0.039	0.934 ± 0.059	0.934 ± 0.100
	10	0.923 ± 0.040	0.933 ± 0.021	0.932 ± 0.044	0.930 ± 0.045	0.926 ± 0.047	0.931 ± 0.061	0.933 ± 0.017
WPBC	0	0.763 ± 0.030	0.763 ± 0.030	0.763 ± 0.030	0.763 ± 0.030	0.763 ± 0.030	0.763 ± 0.030	0.763 ± 0.030
	5	0.711 ± 0.027	0.716 ± 0.015	0.721 ± 0.052	0.734 ± 0.024	0.733 ± 0.068	0.762 ± 0.080	0.734 ± 0.087
	10	0.705 ± 0.082	0.714 ± 0.109	0.720 ± 0.076	0.724 ± 0.064	0.730 ± 0.099	0.759 ± 0.055	0.730 ± 0.020
Sonar	0	0.712 ± 0.072	0.725 ± 0.071	0.716 ± 0.089	0.721 ± 0.110	0.721 ± 0.110	0.722 ± 0.072	0.733 ± 0.055
	5	0.711 ± 0.069	0.719 ± 0.101	0.714 ± 0.076	0.717 ± 0.119	0.716 ± 0.119	0.715 ± 0.068	0.728 ± 0.038
	10	0.673 ± 0.026	0.716 ± 0.012	0.712 ± 0.034	0.711 ± 0.097	0.713 ± 0.025	0.713 ± 0.024	0.718 ± 0.073
Average		0.786 ± 0.051	0.801 ± 0.049	0.799 ± 0.055	0.798 ± 0.051	0.798 ± 0.057	0.805 ± 0.057	0.804 ± 0.064

FIGURE 1: Comparison of robustness of different FRS models on KNN.

FIGURE 2: Comparison of robustness of different FRS models on CART.

FIGURE 3: Comparison of robustness of different FRS models on LSVM.

TABLE 5: Friedman statistical coefficients F_F and corresponding critical values.

Classifier	F_F	Critical value ($\alpha = 0.1$)
KNN	7.1268	1.77
CART	16.3475	
LSVM	17.1893

FIGURE 4: CD diagram of different FRS models on KNN based on Bonferroni-Dunn statistical method.

FIGURE 5: CD diagram of different FRS models on CART based on Bonferroni-Dunn statistical method.

FIGURE 6: CD diagram of different FRS models on LSVN based on Bonferroni-Dunn statistical method.

β -PFRS, K -trimmed FRS, K -means FRS, K -median FRS, and SFRS model.

6. Conclusions

The development of robust FRS is a hot spot in the theory of FRS, which has some advantages in the feature selection of noise information. In this paper, the nonparametric fuzzy membership degree is defined by fuzzy granular calculation, and a FRS model based on representative samples is proposed. Firstly, the distance between the target sample and each representative sample is calculated by looking for the representative sample of each class, then the fuzzy membership degree of the target sample about all classes is determined, a FRS model based on the representative sample is proposed, and the relevant properties of the model are studied. On this basis, a feature selection algorithm based on FRS of representative samples is designed. Experimental results show that the RS-FRS model and feature selection algorithm proposed in this paper are feasible, effective, and robust for the processing of uncertain information systems with noise information, which expands the application field of FRS and the research on feature selection. However, the paper has some limitations. In this paper, the feature selection is carried out by the lower approximation of RS-FRS model. The lower approximation of the sample reflects the degree of certainty to which the sample belongs to the class, while the upper approximation of the sample reflects the

degree of probability to which the sample belongs to the class. In the future research, it is considered to design feature selection algorithms using both upper approximation and lower approximation.

Data Availability

In this paper, the authors use UCI database which is public and open access to obtain in the website <http://www.ics.uci.edu/mllearn/MLRepository.html>.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Jiangsu Province of China (BK2C0200364), the National Natural Science Foundation of China (62001111), the National Key Research and Development Program of China (2019YFE0113800), and the National Natural Science Foundations of China (82072014 and 81901842).

References

- [1] F.-J. Lin, S.-G. Chen, and C.-W. Hsu, "Intelligent back-stepping control using recurrent feature selection fuzzy neural network for synchronous reluctance motor position servo

- drive system," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 3, pp. 413–427, 2019.
- [2] E. Armengol, D. Boixader, À. García-Cerdaña, and J. Recasens, "T-generable indistinguishability operators and their use for feature selection and classification," *Fuzzy Sets and Systems*, vol. 360, pp. 33–48, 2019.
 - [3] S. Kashef and H. Nezamabadi-pour, "A label-specific multi-label feature selection algorithm based on the pareto dominance concept," *Pattern Recognition*, vol. 88, pp. 654–667, 2019.
 - [4] Z. Manbari, F. AkhlaghianTab, and C. Salavati, "Hybrid fast unsupervised feature selection for high-dimensional data," *Expert Systems with Applications*, vol. 124, pp. 97–118, 2019.
 - [5] J. Zhang, J. Miao, Q. Zhao, and Y. Tian, "Multi-task feature selection with sparse regularization to extract common and task-specific features," *Neurocomputing*, vol. 340, pp. 76–89, 2019.
 - [6] F. Nie, S. Yang, R. Zhang, and X. Li, "A general framework for auto-weighted feature selection via global redundancy minimization," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2428–2438, 2019.
 - [7] H. Yuan, J. Li, L. L. Lai, and Y. Y. Tang, "Joint sparse matrix regression and nonnegative spectral analysis for two-dimensional unsupervised feature selection," *Pattern Recognition*, vol. 89, pp. 119–133, 2019.
 - [8] Z. A. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
 - [9] Z. Pawlak and A. Skowron, "Rough sets and boolean reasoning," *Information Sciences*, vol. 177, no. 1, pp. 41–73, 2007.
 - [10] Z. Pawlak and A. Skowron, "Rough sets: some extensions," *Information Sciences*, vol. 177, no. 1, pp. 28–40, 2007.
 - [11] Y. Lin, Q. Hu, J. Liu, J. Chen, and J. Duan, "Multi-label feature selection based on neighborhood mutual information," *Applied Soft Computing*, vol. 38, pp. 244–256, 2016.
 - [12] Y. Li, S. Wu, Y. Lin, and J. Liu, "Different classes' ratio fuzzy rough set based robust feature selection," *Knowledge-Based Systems*, vol. 120, pp. 74–86, 2017.
 - [13] Y. W. Li, Y. J. Lin, J. H. Liu et al., "Feature selection for multi-label learning based on kernelized fuzzy rough sets," *Neurocomputing*, vol. 318, pp. 217–286, 2018.
 - [14] J. M. F. Salido and S. Murakami, "Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations," *Fuzzy Sets and Systems*, vol. 139, pp. 635–660, 2003.
 - [15] Q. Hu, S. An, and D. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection," *Information Sciences*, vol. 180, no. 22, pp. 4384–4400, 2010.
 - [16] S. An, Q. Hu, D. Yu, and J. Liu, "Soft minimum-enclosing-ball based robust fuzzy rough sets," *Fundamenta Informaticae*, vol. 115, no. 2-3, pp. 189–202, 2012.
 - [17] S. An, Q. Hu, W. Pedrycz, P. Zhu, and E. C. C. Tsang, "Data-distribution-aware fuzzy rough set model and its application to robust classification," *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–13, 2015.
 - [18] Q. Hu, L. Zhang, S. An, D. Zhang, and D. Yu, "On robust fuzzy rough set models," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 4, pp. 636–651, 2012.
 - [19] S. Y. Zhao, E. C. C. Tsang, and D. G. Chen, "The model of fuzzy variable precision rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 17, pp. 451–467, 2009.
 - [20] N. Verbiest, C. Cornelis, and F. Herrera, "OWA-FRPS: a prototype selection method based on ordered weighted average fuzzy rough set theory," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pp. 180–190, Springer, Berlin, Heidelberg, 2013.
 - [21] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.
 - [22] C. Wang, Y. Qi, M. Shao et al., "A fitting model for feature selection with fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 741–753, 2017.
 - [23] X. Hu and N. Cercone, "Learning in relational databases: a rough set approach," *Computational Intelligence*, vol. 11, no. 2, pp. 323–338, 1995.
 - [24] Y. Yao and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Information Sciences*, vol. 179, no. 7, pp. 867–882, 2009.
 - [25] R. Jensen, A. Tuson, and Q. Shen, "Finding rough and fuzzy-rough set reducts with SAT," *Information Sciences*, vol. 255, pp. 100–120, 2014.
 - [26] D. B. Zhang and Y. Leung, "On knowledge reduction in inconsistent decision information systems," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 12, no. 05, pp. 651–672, 2004.
 - [27] D. Chen, S. Zhao, L. Zhang, Y. Yang, and X. Zhang, "Sample pair selection for attribute reduction with rough set," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2080–2093, 2012.
 - [28] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 385–389, 2012.
 - [29] Z. Zhang, W. Chen, C. Liu et al., "Robust feature selection based on fuzzy rough sets with representative sample," *Advanced Data Mining and Applications*, pp. 151–165, 2019.
 - [30] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General Systems*, vol. 17, no. 2-3, pp. 191–209, 1990.
 - [31] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.
 - [32] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA, USA, 1984.
 - [33] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.
 - [34] A. Asuncion and D. J. Newman, *UCI Machine Learning Repository*, [http://www.ics.uci.edu/ml/MLRepository.html], University of California, School of Information and Computer Science, Irvine, CA, USA, 2007.
 - [35] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.
 - [36] O. J. Dunn, "Multiple comparisons among means," *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.