WILEY | Hindawi

Research Article

Kernel Neighborhood Rough Sets Model and Its Application

Kai Zeng (iD)[1] **and Siyuan Jing** (iD)[2]

[1]School of Data Science, Guizhou Institute of Technology, No. 1 Caiguan Road, Guiyang 550003, China
[2]School of Computer Science, Leshan Normal University, Binhe Road, Leshan, Sichuan 614000, China

Correspondence should be addressed to Kai Zeng; zengkailink@sina.com

Rough set theory has been successfully applied to many fields, such as data mining, pattern recognition, and machine learning. Kernel rough sets and neighborhood rough sets are two important models that differ in terms of granulation. The kernel rough sets model, which has fuzziness, is susceptible to noise in the decision system. The neighborhood rough sets model can handle noisy data well but cannot describe the fuzziness of the samples. In this study, we define a novel model called kernel neighborhood rough sets, which integrates the advantages of the neighborhood and kernel models. Moreover, the model is used in the problem of feature selection. The proposed method is tested on the UCI datasets. The results show that our model outperforms classic models.

1. Introduction

Rough set theory, which was proposed by Pawlak in 1982, is a powerful mathematical method to study incomplete and imprecise information. This theory has been successfully applied to many fields, such as data mining, decision-making, pattern recognition, machine learning, and intelligent control [1–4]. Kernel rough sets [5] and neighborhood rough sets [6] are two important models in rough set theory.

Hu innovatively proposed the kernel rough sets model [5, 7]. A Gaussian kernel rough sets model-based feature selection method was discussed in [8]. The information fusion problem of imperfect images has also been studied based on Hu's research [9]. Ghosh et al. proposed an efficient Gaussian kernel-based fuzzy rough sets approach for feature selection [10]. A novel fuzzy rough sets model was constructed by combining the hybrid distance and the Gaussian kernel in [11]. A new feature selection method based on kernel fuzzy rough sets and a memetic algorithm were proposed for the transient stability assessment of power systems [12]. In these studies, the information granules are constructed in the kernel structure. The "min" and "max" aggregation operations are used in approximation calculations [13–15]. That is, the decision for a sample is dependent on the nearest sample [7]. The computation of the lower approximation comes with risks if there is noise in the datasets [16]. Data noise can lead to an increase in the classification error rate by using the kernel rough sets model [16].

Neighborhood is an important concept in classification and clustering. To formulate the notion of approximation, the neighborhood system was introduced into the relational model by Lin [17–19]. Yao presented a framework for the formulation, interpretation, and comparison of neighborhood systems and rough sets approximations [20]. Hu et al. investigated the issue of heterogeneous feature subset selection based on neighborhood rough sets [6, 21]. Based on neighborhood granulation, samples are constructed as a family of neighborhood granules to approximate the object sets. The neighborhood model can handle noisy data well based on the tolerance neighborhood relation and probabilistic theory [22]. However, the main limitation of this model is that it cannot describe the fuzziness of samples [16].

Overall, the kernel rough sets model, which has fuzziness, is susceptible to noise in the decision system. The neighborhood rough sets model can handle noisy data but cannot describe the fuzziness of samples. That is, we can construct a new model by the combination of advantage of kernel and neighborhood rough sets.

On the other hand, increasing amounts of high-dimensional data must be processed for some real applications. Currently,

feature selection plays an important role in machine learning and data mining. Neighborhood rough sets and kernel rough sets are widely used in feature selection [23–26]. We also can deal with the feature selection problem by using the new rough sets model.

Based on the motivations above, the contributions of this paper include the following: (1) We define a novel model, the kernel neighborhood rough sets model, which integrates the advantages of the neighborhood and kernel models. (2) Moreover, the model is used in the problem of feature selection. (3) The proposed method is tested on the UCI datasets. The results show that our model yields a better performance than classic models.

This paper is organized as follows. In Section 2, some basic concepts regarding neighborhood rough sets and kernel rough sets are briefly reviewed. In Section 3, the kernel neighborhood rough sets (KNRS) model is investigated in detail. Section 4 shows the application of KNRS to feature evaluation and feature selection. Numerical experiments are reported in Section 5. Finally, Section 6 concludes the paper.

2. Preliminary Knowledge

In this section, we review the kernel rough sets (KRS) model [5] and the neighborhood rough sets (NRS) model [6].

2.1. Kernel Rough Sets (KRS) Model

Definition 1. Suppose U is a nonempty finite set of objects and k is a Gaussian kernel function $k(x_i, x_j) = \exp(-|x_i - x_j|^2/2\delta^2)$, where $|x_i - x_j|^2$ is the Euclidean distance. Therefore, (U, A, k) is a kernel approximation space, where

(1) $k(x_i, x_i) = 1$,

(2) $k(x_i, x_j) = k(x_j, x_i)$,

(3) $k(x_i, x_j) \in [0, 1]$.

Definition 2. Given a kernel approximation space (U, A, k), $X \in F(U)$ is a fuzzy subset of U, and we define the lower and upper approximations of X on the space (U, A, k) as follows:

$$
\begin{aligned}
\underline{K}X(x) &= \inf_{y \in U} \max \left(1 - k(x, y), X(y)\right), \\
\bar{K}X(x) &= \sup_{y \in U} \min \left(k(x, y), X(y)\right).
\end{aligned}
\tag{1}
$$

2.2. Neighborhood Rough Sets (NRS) Model.

Pawlak rough sets theory works only on data sets with discrete attributes [27]. Hu et al. introduced a neighborhood rough sets model for heterogeneous data to avoid discretization. The definition is as follows [6, 28].

Definition 3. Suppose U is a nonempty finite set of objects and Δ is a given distance function. A is a set of features. Therefore, (U, A, δ) is a neighborhood approximation space, where

(1) $\Delta(x_i, x_j) \geq 0$, if and only if $x_i = x_j$, $\Delta(x_i, x_j) = 0$,

(2) $\Delta(x_i, x_j) = \Delta(x_j, x_i)$,

(3) $\Delta(x_i, x_k) \leq \Delta(x_i, x_j) + \Delta(x_j, x_k)$.

Definition 4. Given a neighborhood approximation space (U, A, δ), $\forall x \in U$, and $\delta \geq 0$, $\delta(x)$ is a δ neighborhood of x whose center is x and the radius is δ, where $\delta(x) = \{x \mid \Delta(x, y) \leq \delta, x \in U\}$.
Here, $\delta(x)$ can be considered to be the neighborhood granule.

Remark 1. Given two points $x_i = \{x_i^1, x_i^2, \ldots, x_i^n\}$ and $x_j = \{x_j^1, x_j^2, \ldots, x_j^n\}$ in N-dimensional Euclidean space, the distance between them can be computed as

$$
\Delta(x_i, x_j) = \sqrt{\sum_{l=1}^{N} d_{a_l}(x_i, x_j)^2},
$$

$$
d_{a_l}(x_i, x_j) = \begin{cases} \mathrm{nom\_diff}_{a_l}(x_i, x_j) & a_l \text{ is a nominal attribute,} \\ \mathrm{num\_diff}_{a_l}(x_i, x_j) & a_l \text{ is a numerical attribute,} \end{cases}
\tag{2}
$$

where $\mathrm{nom\_diff}_{a_l}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i^l = x_j^l, \\ 1 & \text{if } x_i^l \neq x_j^l, \end{cases}$ and $\mathrm{num\_diff}_{a_l}(x_i, x_j) = |x_i^l - x_j^l|/\max(V_{a_l}) - \min(V_{a_l})$.

Definition 5. Given a neighborhood approximation space (U, A, δ), for any subset $X \subseteq U$, we define the lower and upper approximations of X on the space (U, A, δ), respectively, as follows:

$$
\begin{aligned}
\underline{N}(X) &= \{x \in X \mid \delta(x) \subseteq X\}, \\
\bar{N}(X) &= \{x \in X \mid \delta(x) \bigcap X \neq \varnothing\}.
\end{aligned}
\tag{3}
$$

The definitions of the lower and upper approximations are the most important concepts in KRS and NRS.

3. Kernel Neighborhood Rough Sets (KNRS) Model

In this section, we study the KNRS model. The definitions and theorems of KNRS are discussed in detail. The kernel neighborhood decision system is also investigated.

3.1. Kernel Neighborhood Rough Sets

Definition 6. Given a kernel neighborhood approximation space (U, A, k, δ), $\delta \geq 0$, where k is a Gaussian kernel function, $\forall x, y \in U$; thus, $[x]_k^\delta$ is a kernel neighborhood granule of x, where

$$
[x]_k^\delta = \begin{cases} 0 & k(x, y) < \delta, \\ k(x, y) & k(x, y) \geq \delta. \end{cases}
\tag{4}
$$

Definition 7. Given a kernel neighborhood approximation space (U, A, k, δ), $\delta \geq 0$, where k is a kernel function, for any fuzzy subset $X \subseteq F(U)$; we define the lower and upper approximations of X on the space (U, A, k, δ), respectively, as follows:

$$
\begin{aligned}
\underline{N}(X) &= \left\{ x_i \in X \,\middle|\, [x]_k^\delta \subseteq X \right\}, \\
\bar{N}(X) &= \left\{ x_i \in X \,\middle|\, [x]_k^\delta \bigcap X \neq \varnothing \right\}.
\end{aligned}
\tag{5}
$$

The method defined above is crisp and has no noise tolerance ability. Here, we propose an improved model that is called variable precision lower and upper approximation.

Definition 8. Given a kernel neighborhood approximation space (U, A, k, δ), $\delta \geq 0$, where k is a kernel function, for any fuzzy subset $X \subseteq F(U)$; the variable precision lower and upper approximations of X are defined as follows, where $|^*|$ denotes the cardinality of the specified set:

$$
\begin{aligned}
\underline{KN}^\alpha(X) &= \left\{ x_i \in X \,\middle|\, \frac{\left|[x]_k^\delta \bigcap X\right|}{\left|[x]_k^\delta\right|} \geq \alpha \right\}, \\
\overline{KN}^\alpha(X) &= \left\{ x_i \in X \,\middle|\, \frac{\left|[x]_k^\delta \bigcap X\right|}{\left|[x]_k^\delta\right|} > 1 - \alpha \right\}.
\end{aligned}
\tag{6}
$$

Then, $0.5 \leq \alpha \leq 1$ as in [22].

Example 1. Consider a kernel neighborhood approximation space (U, A, k, δ), where $U = \{u_1, u_2, u_3, u_4, u_5\}$, $A = \{a_1, a_2, a_3\}$, $\delta = 0.5$, $\alpha = 0.9$, and k is a kernel function. The details are listed in Table 1.
The relation matrix is presented in Table 2. Each line is the kernel neighborhood granule of u_i.

Given a fuzzy set $X = \{0.9/u_1, 0.8/u_2, 0.9/u_3, 0.3/u_4, 0.1/u_5\}$, we obtain $\underline{KN}^\alpha(X) = \overline{KN}^\alpha(X) = \{u_1, u_2, u_3\}$.

Definition 9. Given a kernel neighborhood approximation space (U, A, k, δ), where $\delta \geq 0$ and k is a kernel function, for any fuzzy subset $X \subseteq F(U)$, the positive, negative, and boundary regions of X in the space (U, A, k, δ) are, respectively, expressed as follows:

$$
\begin{aligned}
POS(X) &= \underline{KN}(X), \\
NEG(X) &= U - \overline{KN}(X), \\
BN(X) &= \overline{KN}(X) - \underline{KN}(X).
\end{aligned}
\tag{7}
$$

Theorem 1. *Given a kernel neighborhood approximation space (U, A, k, δ), where $\delta_1 \leq \delta_2 \leq 0$, then*

(1) $[x]_k^{\delta_1} \supseteq [x]_k^{\delta_2}$,

(2) $\underline{KN}^{\delta_1}(X) \subseteq \underline{KN}^{\delta_2}(X)$,

(3) $\overline{KN}^{\delta_1}(X) \subseteq \overline{KN}^{\delta_2}(X)$.

TABLE 1: Approximation space.

ID	a_1	a_2	a_3
u_1	0.1	0.2	0.3
u_2	0.2	0.2	0.3
u_3	0.3	0.1	0.1
u_4	0.8	0.7	0.8
u_5	0.9	0.9	0.7

TABLE 2: Relation matrix.

ID	u_1	u_2	u_3	u_4	u_5
u_1	1	0.90	0.74	0	0
u_2	0.90	1	0.78	0	0
u_3	0.74	0.78	1	0	0
u_4	0	0	0	1	0.88
u_5	0	0	0	0.88	1

Proof. If $\delta_1 \leq \delta_2 \leq 0$, then $[x]_k^{\delta_1} \supseteq [x]_k^{\delta_2}$. Assuming that $[x]_k^{\delta_1} \subseteq X$, then $[x]_k^{\delta_2} \subseteq X$. That is, $x \in \underline{KN}^{\delta_2}(X)$ if $x \in \underline{KN}^{\delta_1}(X)$ but not vice versa. Similarly, we obtain $\overline{KN}^{\delta_1}(X) \subseteq \overline{KN}^{\delta_2}(X)$.

3.2. Kernel Neighborhood Decision System. A kernel neighborhood approximation space is called a variable precision kernel neighborhood decision system which is denoted by $(U, C \cup D, k, \delta, \alpha)$. C and D are condition and decision features, respectively.

Definition 10. Consider a variable precision kernel neighborhood decision system $(U, C \cup D, k, \delta, \alpha)$, where $\delta \geq 0$, $0.5 \leq \alpha \leq 1$, and k is a kernel function. Suppose $B \subseteq C$ and $U/D = \{D_1, D_2, \dots, D_n\}$. We, respectively, define the lower and upper approximations of D with respect to feature subset B as follows:

$$
\begin{aligned}
\underline{KN}_B^\alpha(D) &= \bigcup_{i=1}^n \underline{KN}_B^\alpha(D_i), \\
\overline{KN}_B^\alpha(D) &= \bigcup_{i=1}^n \overline{KN}_B^\alpha(D_i).
\end{aligned}
\tag{8}
$$

Definition 11. Consider a variable precision kernel neighborhood decision system, $(U, C \cup D, k, \delta, \alpha)$, where $\delta \geq 0$, $0.5 \leq \alpha \leq 1$, and k is a kernel function. Suppose $B \subseteq C$ and $U/D = \{D_1, D_2, \dots, D_n\}$. The positive, negative, and boundary regions of D in the space (U, A, k, δ) are, respectively, defined as follows:

$$
\begin{aligned}
POS_B(D) &= \underline{KN}_B(D), \\
NEG_B(D) &= U - \overline{KN}_B(D), \\
BN_B(D) &= \overline{KN}_B(D) - \underline{KN}_B(D).
\end{aligned}
\tag{9}
$$

The size of the boundary increases the uncertainty in the decision system. The samples in the boundary generally have the same condition features but belong to different decision classes. This discrepancy leads to the poor classification

```
Input: decision system (U, C ∪ D, k, δ, α) and stopping threshold ε
Output: selected features red
1.   red ← ∅
2.   while red ≠ C
3.       for each aᵢ ∈ (C − red)
4.           compute |POS_{red∪aᵢ}(D)|/|U|
5.           compute SIG(aᵢ, red, D) = |POS_{red∪aᵢ}(D)|/|U| − |POS_{red}(D)|/|U|
6.       end for
7.   find the maximal significance and the corresponding attribute aᵢ
8.       if SIG(aᵢ, red, D) > ε
9.           red ← red ∪ aᵢ
10.      else
11.          exit while
12.      end if
13.  end while
14.  return red
```

ALGORITHM 1: Feature selection based on a kernel neighborhood rough set (FSKNRS).

TABLE 3: Data description [3].

ID	Data	Samples	Features	Class
1	Wine	178	13	3
2	wdbc	569	31	2
3	Heart	270	13	2
4	Glass	214	9	7
5	Anneal	798	38	6

performance of the employed algorithms. We should select a proper feature subset that increases the size of the positive region and decreases the size of the boundary region.

Definition 12. Consider a variable precision kernel neighborhood decision system $(U, C \cup D, k, \delta, \alpha)$, where $\delta \geq 0$, $0.5 \leq \alpha \leq 1$, and k is a kernel function. Suppose $B \subseteq C$ and $U/D = \{D_1, D_2, \ldots, D_n\}$. The dependency degree of D relative to B is defined as follows:

$$\gamma_B = \frac{|POS_B(D)|}{|U|}. \tag{10}$$

The dependency degree reflects the approximating power of a condition feature set. A higher approximation degree means that the samples that are described by feature subset B are more consistent with decision D.

Theorem 2. *Given a variable precision kernel neighborhood decision system* $(U, C \cup D, k, \delta, \alpha)$, *where* $C_1 \subseteq C_2 \subseteq C$, *then*

(1) $[x]_k^{\delta,C_1} \supseteq [x]_k^{\delta,C_2}$,

(2) $\underline{KN}_{C_1}(D) \subseteq \underline{KN}_{C_2}(D)$,

(3) $POS_{C_1}(D) \subseteq POS_{C_2}(D)$.

Proof. If $\delta_1 \subseteq \delta_2 \subseteq 0$, then $[x]_k^{\delta,C_1} \supseteq [x]_k^{\delta,C_2}$. Assuming that $[x]_k^{\delta,C_1} \subseteq D$, then $[x]_k^{\delta,C_2} \subseteq D$. That is, $x \in \underline{KN}_{C_2}(D)$ if $x \in \underline{KN}_{C_1}(D)$ but not vice versa. Similarly, we obtain $POS_{C_1}(D) \subseteq POS_{C_2}(D)$.
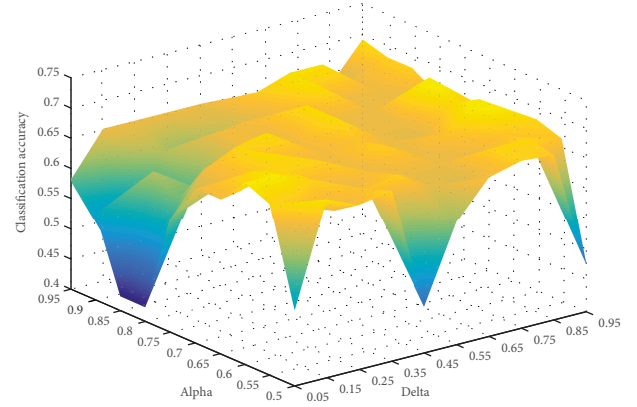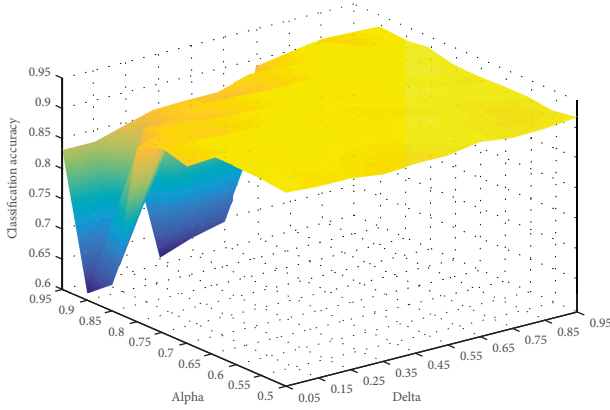
4. Feature Selection Based on KNRS

One of the most important applications of the information rough sets theory is the evaluation of the classification power of the attributes. In this section, we define the significance of the feature subsets. The feature selection algorithm is also discussed.

Definition 13. Consider a variable precision kernel neighborhood decision system $(U, C \cup D, k, \delta, \alpha)$, where $\delta \geq 0$, $0.5 \leq \alpha \leq 1$, and k is a kernel function. Suppose $B \subseteq C$ and $U/D = \{D_1, D_2, \ldots, D_n\}$. The significance of feature a in B is defined as follows:

$$SIG(a, B, D) = \gamma_B(D) - \gamma_{B-a}(D). \tag{11}$$

$SIG(a, B, D)$ is used to evaluate the significance of attribute a in subset B. That is, a is an important feature if it increases the dependency degree of subset B. As mentioned in Definition 11, we need subset B to be more consistent with decision D. Thus, we define the feature selection algorithm as follows.

It is impractical to obtain the optimal subset of features from $2^n - 1$ candidates through an exhaustive search, where n is the number of features. We use a forward greedy search algorithm, which is usually more efficient than a standard brute-force exhaustive search [4]. That is, one starts with an empty set of attributes and adds features to the subset of selected attributes one by one. Each selected attribute maximizes the increment of significance of the current subset.
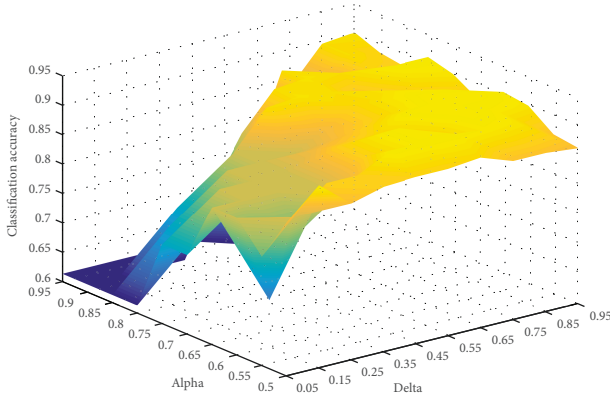
FIGURE 1: Classification accuracy with δ and α on wine.

FIGURE 2: Classification accuracy with δ and α on wdbc.

FIGURE 3: Classification accuracy with δ and α on glass.
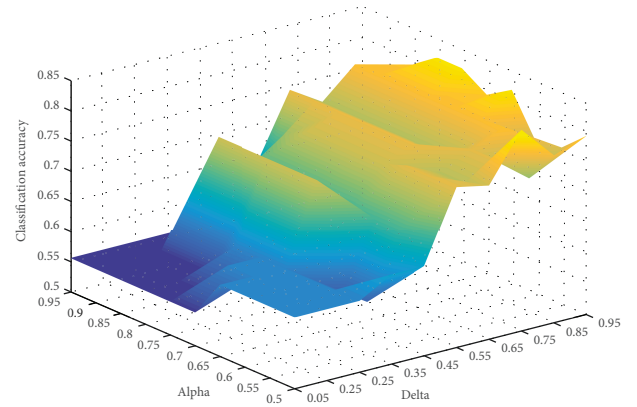
FIGURE 4: Classification accuracy with δ and α on heart.

Corollary 1. *Given a variable precision kernel neighborhood decision system* $(U, C \cup D, k, \delta, \alpha)$, *where* $C_1 \subseteq C_2 \subseteq C$, *if* $x \in POS_{C_1}(D)$, *then* $x \in POS_{C_2}(D)$.

Proof. Please refer to the proof of Theorem 2.

Corollary 1 shows that an object must belong to the positive region with respect to the feature sets (such as C_2) if the object belongs to the positive region with respect to the feature subset (such as C_1, where $C_1 \subseteq C_2 \subseteq C$). Therefore, it is unnecessary to consider every object when computing the positive region. Then, we can obtain a fast feature selection algorithm by improving Algorithm 1.

5. Experimental Analysis

In this section, we evaluate the effectiveness of KNRS through a series of experiments. The data sets are downloaded from the UCI machine learning repository (http://archive.ics.uci.edu/ml/index.php) as [3] and are described in Table 3. The numerical attributes of the samples are linearly normalized as follows:

$$x = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}, \tag{12}$$

where $x_{\min}$ and $x_{\max}$ are the bounds of the given attribute. Two popular machine leaning algorithms, namely, CART
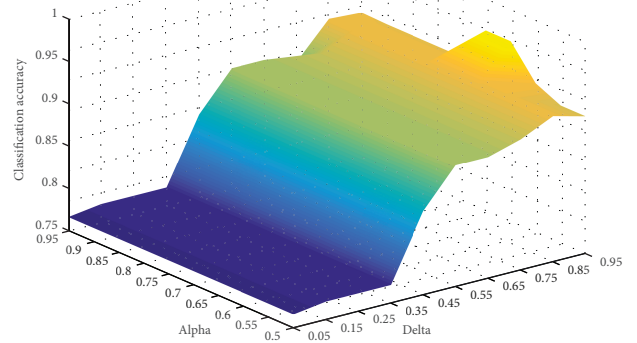
FIGURE 5: Classification accuracy with δ and α on anneal.

and RBF SVM, are used to evaluate the quality of the selected features. The experiments are run in a 10-fold cross-validation mode. The parameters of CART and RBF SVM are set to the default values (by using MATLAB toolkit osu_svm3.00).

5.1. Discussion of the Parameters. The neighborhood metric δ and precision degree α are two key parameters that affect the feature selection results. Here, δ and α are analyzed through

```
Input: decision system (U, C ∪ D, k, δ, α) and stopping threshold ε
Output: selected features red
1.    red ← ∅, S ← U
2.    while red ≠ C
3.        for each a_i ∈ (C − red)
4.            for each object o_j ∈ S
5.                if o_j ∈ POS(D)
6.                    S ← S − o_j
7.            end for
8.            compute |POS_{red∪a_i}(D)|/|U| by the object subset U − S
9.            compute SIG(a_i, red, D) = |POS_{red∪a_i}(D)|/|U| − |POS_{red}(D)|/|U|
10.       end for
11.       find the maximal significance and the corresponding attribute a_i
12.       if SIG(a_i, red, D) > ε
13.           red ← red ∪ a_i
14.       else
15.           exit while
16.       end if
17.   end while
18.   return red
```

ALGORITHM 2: Fast feature selection based on a kernel neighborhood rough set (FFSKNRS).

a series of experiments. We set δ from 0.05 to 0.95 with a step size of 0.1. We set the precision degree α from 0.5 to 0.95 with a step size of 0.05. The evaluation criterion is the classification accuracy in the selected feature subset with parameters δ and α.

According to Figures 1–5, on most of the data sets, higher precision is achieved in a larger area when δ is between 0.5 and 0.95 and α is between 0.5 and 0.75. That is, the FNRS model is feasible and stable in most cases.

5.2. Effectiveness of the Fast Feature Selection. We propose feature selection based on a kernel neighborhood rough set (Algorithm 1: FSKNRS) and a fast version (Algorithm 2: FFSKNRS). We evaluated the run times of the two methods. The results are listed in Table 4. The fast feature selection algorithm that is based on KNRS yields a better performance than Algorithm 1 is.

5.3. Comparison of the Effectiveness in Feature Selection. In this section, we select KRS [5], NRS [22], and neighborhood entropy (NE) [28] as the comparison models with KNRS. The feature subsets that are selected by different algorithms are presented in Table 5. The features are presented in the order in which they were added to the feature space.

The KNS model cannot obtain any subsets of the data set "glass." Most of the feature subsets are slightly different. A small difference in feature quality may lead to a completely different ranking. The orders of the feature subsets reflect the relative significance of the features in terms of the corresponding measures. Therefore, the large differences among these selected features are due to the differences in the qualities of the feature measures.

Then, we build classification models with the selected features and test their classification performance based on a

TABLE 4: Comparison of run times (sec).

Data	Algorithm 1: FSKNRS	Algorithm 2: FFSKNRS
Wine	0.27	0.15
wdbc	10.4	2.43
Glass	0.31	0.23
Heart	0.64	0.21
Anneal	13.01	3.33
Average	*4.93*	*1.27*

10-fold cross-validation. The average value and standard deviation are used to measure the classification performance. We compare KNRS, KRS, NRS, and NE in Tables 6 and 7, where the learning algorithms of CART and RBF SVM are used to evaluate the selected features.

The comprehensive correlation results are shown in Table 8. The number of wins for KNRS, KRS, NRS, and NE are 4, 0, 4, and 2, respectively. KNRS achieves the highest average classification accuracy by using fewer features. It is thus concluded that KNRS outperforms the other feature measures. We can interpret the results from two aspects. For the kernel model (KRS), the lower approximation is computed by the "min" operation. Then, the decision on a sample is dependent on the nearest sample. This procedure can lead to a decision error if the nearest sample is a noise point. We obtain a lower classification accuracy when there is considerable data noise in the samples. In contrast, neighborhood-based binary relations can only be expressed in terms of 0 or 1. Neighborhood models (such as NRS and NE) cannot describe the fuzziness delicately like the kernel model. KNRS is a better choice because it integrates the advantages of the kernel and neighborhood models.

Table 5: Subsets of the features selected with KNRS, KRS, NRS, and NE.

Data	KNRS	KRS	NRS	NE
Wine	12, 13, 2, 1	13, 10, 7, 1, 5, 2, 12	13, 10, 7, 6, 1	7, 1, 11, 4
wdbc	29, 22, 23, 26	24, 29, 23, 13, 22, 26, 10	9, 23, 22, 26, 31, 13	24, 29, 23
Glass	3, 9, 4, 5, 1, 2, 7	N/A	8, 4, 7, 5, 9, 3, 1, 2	3, 7, 4, 9, 5, 1
Heart	3, 13, 12, 11, 2	2, 3,12, 11, 13, 7, 1	10, 2, 8, 4, 1, 3, 5	13, 12, 3, 10, 1, 4, 5
Anneal	12, 3, 8, 5	5, 12, 7, 3, 1	12, 5, 3, 7, 9, 13	12, 3, 8, 5
Average	*4.8*	*6.5*	*6.4*	*4.8*

Table 6: Classification accuracies based on CART (%).

Data	KNRS	KRS	NRS	NE
Wine	90.00 ± 8.20	87.50 ± 9.73	91.11 ± 8.76 (*)	89.93 ± 8.57
wdbc	92.11 ± 3.60	91.93 ± 3.22	93.84 ± 4.43	93.85 ± 2.38 (*)
Glass	66.93 ± 11.44 (*)	0	66.90 ± 11.44	65.13 ± 14.65
Heart	81.85 ± 7.90 (*)	75.93 ± 5.01	68.52 ± 7.25	80.00 ± 7.65
Anneal	95.55 ± 1.82	96.44 ± 1.27	98.99 ± 0.82 (*)	95.55 ± 1.82
Average	*85.288*	*70.36*	*83.872*	*84.892*

Table 7: Classification accuracies based on RBF SVM (%).

Data	KNRS	KRS	NRS	NE
Wine	97.78 ± 3.88 (*)	97.7 ± 3.88	97.22 ± 3.93	94.93 ± 4.12
wdbc	96.85 ± 2.70	96.84 ± 1.81	97.20 ± 1.69 (*)	96.66 ± 1.75
Glass	67.86 ± 11.70 (*)	0	67.8 ± 11.70	64.08 ± 11.22
Heart	82.22 ± 7.96	81.11 ± 8.98	77.78 ± 10.33	84.07 ± 7.82 (*)
Anneal	94.77 ± 2.22	96.33 ± 1.18	98.66 ± 0.71 (*)	94.77 ± 2.22
Average	*87.896*	*74.396*	*87.732*	*86.902*

Table 8: Comprehensive correlations of the models.

Data	Win	Average accuracy	Average number of features
KNRS	4	86.59	4.8
KRS	0	72.37	6.5
NRS	4	85.80	6.4
NE	2	85.89	4.8

6. Conclusion and Future Work

As we know that, genetic algorithms and neural networks are famous mathematical model for pattern recognition, machine learning, and intelligent control. However, rough sets theory has also been successfully applied to these fields [5, 13, 15]. In this study, we define a novel model, the kernel neighborhood rough sets model, which integrates the advantages of the neighborhood and kernel models. Moreover, the model is used in the problem of feature selection. The parameters of KNRS are also discussed in detail. Then, we evaluate the effectiveness of the fast feature selection algorithm. A comparison of the results shows that our model yields a better performance than classic models.

There are two potential directions for future work. First, many other rough sets models can be incorporated into KNRS, such as fuzzy rough sets and Pawlak rough sets. Evaluating the significance of features by using the confluent modes is an important issue. Second, the application of our model to big data is necessary. Consequently, the development of a version of KNRS within a distributed framework requires further attention.

Data Availability

The experimental data sets can be downloaded from the UCI machine learning repository (http://archive.ics.uci.edu/ml/index.php).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] K. Zeng, "Preference mining using neighborhood rough set model on two universes," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 6975458, 13 pages, 2016.

[2] K. Zeng and K. She, "Variable precision neighborhood rough sets on two universes," in *2013 IEEE International Conference on Granular Computing (GrC)*, pp. 418–422, Beijing, China, December 2013.

[3] K. Zeng, K. She, and X. Z. Niu, "Multi-granulation entropy and its applications," *Entropy*, vol. 15, no. 12, pp. 2288–2302, 2013.

[4] K. Zeng, K. She, and X. Z. Niu, "Feature selection with neighborhood entropy-based cooperative game theory," *Computational Intelligence and Neuroscience*, vol. 2014, Article ID 479289, 10 pages, 2014.

[5] Q. H. Hu, L. Zhang, D. G. Chen, W. Pedrycz, and D. R. Yu, "Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications," *International Journal of Approximate Reasoning*, vol. 51, no. 4, pp. 453–471, 2010.

[6] Q. H. Hu, D. R. Yu, J. F. Liu, and C. X. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[7] Q. Hu, D. Yu, W. Pedrycz, and D. Chen, "Kernelized fuzzy rough sets and their applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1649–1667, 2010.

[8] D. Chen, Q. Hu, and Y. Yang, "Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets," *Information Sciences*, vol. 181, no. 23, pp. 5169–5179, 2011.

[9] Q. Shi, W. Chen, Q. Qin, and G. Ma, "Gaussian kernel-based fuzzy rough set for information fusion of imperfect images," in *2014 12th International Conference on Signal Processing (ICSP)*, pp. 900–904, Hangzhou, China, October 2014.

[10] S. Ghosh, P. S. V. S. S. Prasad, and C. R. Rao, "An efficient Gaussian kernel based fuzzy-rough set approach for feature selection," in *Multi-disciplinary Trends in Artificial Intelligence. MIWAI 2016*, Lecture Notes in Computer Science, C. Sombattheera, F. Stolzenburg, F. Lin, and A. Nayak, Eds., pp. 38–49, Springer, Cham, 2016.

[11] A. Zeng, T. Li, D. Liu, J. Zhang, and H. Chen, "A fuzzy rough set approach for incremental feature selection on hybrid information systems," *Fuzzy Sets and Systems*, vol. 258, no. 1, pp. 39–60, 2015.

[12] X. Gu, Y. Li, and J. Jia, "Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 64, no. 64, pp. 664–670, 2015.

[13] C. Wang, Q. He, M. Shao, and Q. Hu, "Feature selection based on maximal neighborhood discernibility," *International Journal of Machine Learning and Cybernetics*, pp. 1–12, 2017.

[14] C. Wang, Q. He, M. Shao, Y. Xu, and Q. Hu, "A unified information measure for general binary relations," *Knowledge-Based Systems*, vol. 135, pp. 18–28, 2017.

[15] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 2986–2999, 2017.

[16] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowledge-Based Systems*, vol. 111, pp. 173–179, 2016.

[17] T. Y. Lin, "Neighborhood systems and approximation inrelational databases and knowlege bases," in *Proceedings of the Fourth International Symposium on Methodologies of Intelligent Systems, Poster Session*, pp. 75–86, October 1989.

[18] T. Y. Lin, "Granular computing: from rough sets and neighborhood systems to information granulation and computing with words," in *European Congress on Intelligent Techniques and Soft Computing*, 1997.

[19] T. Y. Lin, "Granular computing on binary relations in data mining and neighborhood systems," *Rough Sets in Knowledge Discovery*, vol. 1, pp. 107–121, 1998.

[20] Y. Y. Yao, "Relational interpretations of neighborhood operators and rough set approximation operators," *Information Sciences*, vol. 111, no. 1–4, pp. 239–259, 1998.

[21] G. P. Lin, Y. H. Qian, and J. J. Li, "NMGRS: neighborhood-based multigranulation rough sets," *International Journal of Approximate Reasoning*, vol. 53, no. 7, pp. 1080–1093, 2012.

[22] S. Y. Jing, K. She, and S. Ali, "A universal neighbourhood rough sets model for knowledge discovering from incomplete heterogeneous data," *Expert Systems*, vol. 30, no. 1, pp. 89–96, 2013.

[23] 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.

[24] Y. Yang, D. Chen, H. Wang, and X. Wang, "Incremental perspective for feature selection based on fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1257–1273, 2018.

[25] A. K. Das, S. Das, and A. Ghosh, "Ensemble feature selection using bi-objective genetic algorithm," *Knowledge-Based Systems*, vol. 123, pp. 116–127, 2017.

[26] J. Dai, Q. Hu, H. Hu, and D. Huang, "Neighbor inconsistent pair selection for attribute reduction by rough set approach," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 937–950, 2018.

[27] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[28] Q. H. Hu and D. R. Yu, "Neighborhood entropy," in *2009 International Conference on Machine Learning and Cybernetics*, vol. 3, pp. 1776–1782, Hebei, China, July 2009.