Research Article

Feature Subset Selection Based on Variable Precision Neighborhood Rough Sets

Yingyue Chen[1], Yumin Chen[2],*,

[1]*School of Economics and Management, Xiamen University of Technology, Xiamen, 361024, China*
[2]*School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, 361024, China*

ABSTRACT

Rough sets have been widely used in the fields of machine learning and feature selection. However, the classical rough sets have the problems of difficultly dealing with real-value data and weakly fault tolerance. In this paper, by introducing a neighborhood rough set model, the values of decision systems are granulated into some condition and decision neighborhood granules. A concept of neighborhood granular swarm is defined in a decision system. Then the sizes of a neighborhood granule and a neighborhood granular swarm are also given. In order to enhance the fault-tolerant ability of classification systems, we define some concepts of granule inclusion, variable precision neighborhood approximation sets and positive region. We propose a variable precision neighborhood rough set model, and analyze its property. Furthermore, based on the positive region of a variable precision neighborhood, we give the significance of an attribute and use it to select feature subsets. A feature subset selection algorithm to the variable precision neighborhood rough sets is designed. Finally, the feature selection algorithm is carried out on the UCI datasets, and the selected features are tested by the support vector machine (SVM) classification algorithm. Theoretical analysis and experiments show that the proposed method can find the effective and compact feature subsets, which have abilities of fault tolerance.

1. INTRODUCTION

The classical Rough Set Theory (RST) proposed by Pawlak [1] refers to the whole study objects as a domain. An equivalence relation is used to granulate the domain into several exclusive equivalence classes, called information granules, which are employed in describing the arbitrary concepts in the domain. At present, the RST has been widely used in the fields of attribute reduction [2–5], feature selection [6–9], gene selection [10,11] and classification [12–15], etc. However, as an effective granular computing model, the RST is only suitable for dealing with discrete data. It cannot directly process continuous data, which seriously affects the areas of practical applications. For the disposing of continuous data, we can discretize the continuous data. But this transformation may result to an information loss, and the classification accuracy largely depends on a particular discretization method. In order to solve the problem, the neighborhood RST [16–18] is introduced to granulate these continuous data and some neighborhood granules are constructed for the convenient analyzing and processing real-value data. The neighborhood RST is intuitive, which can directly handle continuous or real-value data without discretization. Compared with classical RST, it eliminates the discretization process and extends an application scope of the Pawlak RST.

Yao [19] initially proposed the neighborhood RST and Hu *et al.* [20] developed its applications in a classification field. It has been widely

used in attribute reduction [21–24], feature selection and extraction [25–27], classification and clustering [28–30], gene selection [31–33], image processing [34,35] and other fields. The neighborhood RST uses the concept of neighborhood positive region. The objects, that their condition granules are completely contained in their decision granules, form the neighborhood positive region. Different neighborhood parameters make the neighborhood positive regions different. The smaller of the neighborhood parameter and the finer neighborhood granule is. Then the neighborhood positive region becomes larger. The neighborhood positive region also relates to the number of attributes. While attributes are gradually increased, the neighborhood granules become smaller. Then the neighborhood positive region becomes larger. If the neighborhood parameter of a neighborhood classification system is fixed, then the neighborhood positive region is also unchanged. The attribute reduction in a neighborhood system refers to the deletion of redundant attributes in the case of preserving the same neighborhood positive region. However, when some noise are added to a neighborhood system, the neighborhood positive region is changed, which affects the accuracy of a classification. For example, a condition granule is completely contained in a decision granule, and when a noise is added, the condition granule may be partially included in the decision granule. Therefore, it is necessary to consider the circumstance of granule partial-inclusion under a noisy condition. The variable precision RST proposed by Ziarko [36], is a rough set model with fault-tolerant ability. It can overcome the limitation of classical RST with the hard-to-process noisy data. It is an

informational filtering, which persists in some correct information with a great extent and ignores the noisy information. The classification precision can be controlled by a variable factor of the positive region. The variable precision RST has been widely used in many fields such as decision analysis [37], process control [38], data clustering [39] and attribute selection [40].

Feature selection [41–45] also called feature subset selection or attribute selection, refers to selecting N features from the existing M features to obtain the optimization of classification accuracy in the knowledge classification system. Selecting effective features from original features to reduce the dimension of a dataset is an important way to improve the performance of learning algorithms and is also the key data preprocessing step in pattern recognition. RST is a good feature selection tool [46–48], which goal is using rough set positive region to construct classification accuracy. Classical variable precision RST is suitable to feature selection of noisy data. However, the classical variable precision RST is mainly applied to the feature selection based on equivalence relation. While the neighborhood relation is not a strict equivalence relation, the classical variable precision RST is not suitable for neighborhood knowledge classification systems. Therefore, we use the neighborhood RST to granulate the real-value data and construct two types of granules, which are neighborhood granules of condition features and equivalent granules of decision features. Furthermore, we define a novel concept of granule inclusion degree that the condition granules are partly contained in the decision granules, which can robustly deal with noisy datasets. Based on the granule inclusion degree, we construct neighborhood variable precision positive region sets and the significance of a feature. Permitting a certain degree of error classification rate in a feature selection task, we achieve the optimal feature subset. A feature selection algorithm based on variable precision neighborhood granulation is further designed, and some UCI datasets are applied to test the classification performance of the selected feature subsets. The experiments show that the feature subsets selected by the proposed algorithm still have better classification performance for noisy datasets.

The main innovative work of this paper is as follows: (1) In order to enhance the fault-tolerant ability of classification systems, we define some concepts of granule inclusion, variable precision neighborhood dependence and approximation sets. (2) Based on these terminologies, a variable precision neighborhood rough set model is proposed, which is suitable to deal with real-value and noisy data. (3) Furthermore, we prove the monotonicity of variable precision neighborhood dependence with increasing features and a feature subset selection algorithm to the variable precision neighborhood rough sets is designed.

The remainder sections are structured as follows: An introduction to neighborhood granulation of classification systems is presented in Section 2. In Section 3, the granule inclusion degree is introduced. Then we propose a variable precision neighborhood rough set model. In Section 4, we propose a feature selection method based on variable precision neighborhood rough sets and design an algorithm for feature subset selecting. In Section 5, for contributing to understand our proposed method, we carry out some experiments in decision systems. Finally, this paper is concluded with some discussions and remarks in Section 6.

2. NEIGHBORHOOD GRANULATION OF CLASSIFICATION SYSTEMS

The RST proposed by Pawlak, a mathematician of Poland, is one of the most widely used models in classification systems. In a machine learning field, huge amounts of real-value data need to be discretized, but the discretization process easily leads to a loss of classification information. Aiming at the limitation of Pawlak RST, a neighborhood RST is introduced to granulate these real-value data. In RST, an equivalence class is regarded as a granule. As for neighborhood RST, neighborhood granules are constructed.

Definition 1. [20] Let $CS = (U, A, D)$ be a classification system, where $U = \{x_1, x_2, ..., x_n\}$ is a set of samples or objects. $A = \{a_1, a_2, ..., a_m\}$ is a set of condition features or attributes. $D = \{d\}$ represents a decision feature or attribute. The value of a sample in the condition feature set A is a real type of data, and the value of the sample in the decision feature D is a symbolic type or a discrete type data.

Definition 2. [20] Let $CS = (U, A, D)$ be a classification system, for any sample $x, y \in U$, a feature subset $P \subseteq A$, where $P = \{a_1, a_2, ..., a_n\}$, a distance function on the feature subset P is defined as:

$$\Delta_P(x, y) = \left(\sum_{i=1}^{n} (|v(x, a_i) - v(y, a_i)|)^s\right)^{1/s},$$

where $v(x, a_i)$ represents the value of the sample x on a feature a_i. If $s = 1$, it is named Manhattan distance, and if $s = 2$, it is named Euclidean distance.

Definition 3. Let $CS = (U, A, D)$ be a classification system, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$ and a condition feature subset $P \subseteq A$, the δ neighborhood granule of x on P is defined as

$$n_P^\delta(x) = \{y | x, y \in U, \Delta_P(x, y) \leq \delta\},$$

If $G_P^\delta = \{n_P^\delta(x) | \forall x \in U\}$, then G_P^δ is called a neighborhood granular swarm of P.

The values on decision features in classification systems are symbolic data. Therefore, the samples are divided into equivalent classes on the decision features. The neighborhood parameter sets zero, so these samples are granulated into equivalent granules, which are called decision equivalent granules, and they are expressed as

$$n_D^0(x) = \{y | x, y \in U, \Delta_D(x, y) = 0\}.$$

From the above definitions, we can see that a neighborhood granule is degenerated into an equivalent granule when the neighborhood parameter is 0. So the equivalent granule is a special neighborhood granule. The neighborhood granular swarm is a collection of neighborhood granules.

Property 1. The neighborhood granule $n_P^\delta(x)$ of x satisfies the following properties:

1. $n_P^\delta(x) \neq \varnothing$;

2. $x \in n_P^\delta(x)$;

3. $y \in n_P^\delta(x) \leftrightarrow x \in n_P^\delta(y)$;

4. $\cup_{x \in U} n_P^\delta(x) = U$.

Definition 4. Let $CS = (U, A, D)$ be a classification system, $\delta \in [0, 1]$ is a neighborhood parameter, for a feature subset $P \subseteq A$, it determines a neighborhood relation, which is defined as

$$NR_\delta(P) = \{(x, y) \in U \times U | \Delta_P(x, y) \le \delta\}$$

The $U/NR_\delta(P)$ is called a covering of the domain U, which is induced by a neighborhood relation, and the covering is a collection of neighborhood granules. Neighborhood relation is a kind of similarity relation that satisfies reflexive and symmetric properties, rather than equivalence relation. When a neighborhood parameter δ is equal to 0, the neighborhood relation is degenerated into an equivalence relation. So the equivalence relation is a special neighborhood relation. The rough sets based on neighborhood relations can handle the real-value data, while the rough sets based on equivalence relations can only deal with the symbolic data.

Definition 5. Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$ and a feature subset $P \subseteq A$, let $n_P^\delta(x)$ be a neighborhood granule, then the size of granule $n_P^\delta(x)$ is defined as

$$Size(n_P^\delta(x)) = \frac{|n_P^\delta(x)|}{|U|}$$

where $|.|$ represents the cardinality of a set. It is easy to know that the size of neighborhood granule holds the property: $\frac{1}{|U|} \le Size(n_P^\delta(x)) \le 1$.

Definition 6. Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$ and a feature subset $P \subseteq A$, let G_P^δ be a neighborhood granular swarm, then the size of neighborhood granular swarm G_P^δ is defined as

$$GM(G_P^\delta) = \frac{1}{|U|} \sum_{i=1}^{|U|} Size(n_P^\delta(x_i)) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |n_P^\delta(x_i)|$$

It is easy to know that the size of neighborhood granular swarm holds the property: $\frac{1}{|U|} \le GM(G_P^\delta) \le 1$.

Definition 7. [20] Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any subset of samples $X \subseteq U$ and any subset of features $P \subseteq A$, the neighborhood lower and upper approximations of X on P are defined as

$$P_*(X)_\delta = \{x \in U | n_P^\delta(x) \subseteq X\},$$

$$P^*(X)_\delta = \{x \in U | n_P^\delta(x) \cap X \ne \varnothing\}.$$

The order couple $< P_*(X)_\delta, P^*(X)_\delta >$ is called a neighborhood rough set, and it approximates to the X set.

Proposition 1. *Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$ and two feature subsets $P, Q \subseteq A$, suppose $n_P^\delta(x), n_Q^\delta(x)$ are two neighborhood granules of x on P, Q respectively, then we have*

1. *If $P \subseteq Q$ then $n_P^\delta(x) \supseteq n_Q^\delta(x)$.*

2. *If $0 \le \gamma \le \delta \le 1$ then $n_P^\gamma(x) \subseteq n_P^\delta(x)$.*

Proof. 1. Since $P \subseteq Q$, we have $Q = P \cup R$. Let $\forall x \in U$, then $n_P^\delta(x) = \{y | x, y \in U, \Delta_P(x, y) \le \delta\}$ and $n_{P \cup R}^\delta(x) = \{y | x, y \in$

$U, \Delta_{P \cup R}(x, y) \le \delta\}$. According to the definition of distance function, we know $\Delta_P(x, y) \le \Delta_{P \cup R}(x, y)$. Thus, $\Delta_P(x, y) \le \Delta_Q(x, y)$. Therefore, $n_P^\delta(x) \supseteq n_Q^\delta(x)$.

2. For $\forall x \in U$, according to the definition of neighborhood granule, we know $n_P^\gamma(x) = \{y | x, y \in U, \Delta_P(x, y) \le \gamma\}$ and $n_P^\delta(x) = \{y | x, y \in U, \Delta_P(x, y) \le \delta\}$. Since $0 \le \gamma \le \delta \le 1$, we can easily obtain $n_P^\gamma(x) \subseteq n_P^\delta(x)$.

3. VARIABLE PRECISION NEIGHBORHOOD ROUGH SETS

In 1993, Ziarko proposed the variable precision RST [36], which10pt is a generalization of the classical RST and it endures certain classification error rates. However, the variable precision RST can only deal with symbolic or discrete data, which cannot directly handle a large number of symbolic or continuous data in real world. To overcome the deficiency, we introduce neighborhood rough sets, while relaxing the classification accuracy, and propose a variable precision neighborhood rough set model.

Definition 8. Let S, T be two neighborhood granules on domain U, the granule inclusion degree of S contained in T is defined as

$$I(S, T) = \frac{|S \sqsubseteq T|}{|U|} = \frac{|U - (S - T)|}{|U|}.$$

The granule inclusion degree reflects an inclusion relation between two granules. A granule is essentially a set. Figure 1 illustrates a variation for the granule inclusion degree. Suppose S and T are two granules in Figure 1. The light red part of the ellipse represents $S - T$. From the left to right, the size of $S - T$ decreases, so the size of $U - (S - T)$ increases. Finally, the granule inclusion degree increases. It is consistent with that the inclusion degree of S contained in T increases.

Example 1. Suppose a domain $U = \{x_1, x_2, x_3, x_4, x_5\}$. The $S = \{x_1, x_2, x_3\}$ and $T = \{x_1, x_2, x_5\}$ are two neighborhood granules. The granule inclusion degree is $I(S, T) = \frac{|U - (S - T)|}{|U|} = \frac{|\{x_1, x_2, x_4, x_5\}|}{|\{x_1, x_2, x_3, x_4, x_5\}|} = 0.8$.

Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$, a conditional feature subset $P \subseteq A$ and the decision feature D, suppose $n_P^\delta(x)$ is a condition neighborhood granule of x on P, and $n_D^0(x)$ is a decision equivalent granule of x on D. Then we have the granule inclusion degree of the condition granule contained in the decision granule:

$$I(n_P^\delta(x), n_D^0(x)) = \frac{|n_P^\delta(x) \sqsubseteq n_D^0(x)|}{|U|} = \frac{|U - (n_P^\delta(x) - n_D^0(x))|}{|U|}$$

Proposition 2. *Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$ and two*
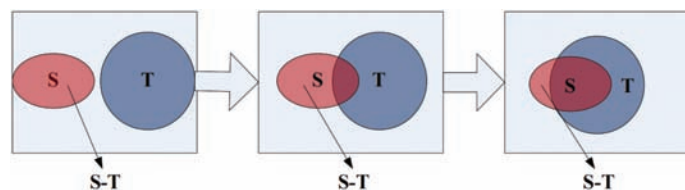
Figure 1 | Illustration for the granule inclusion degree.

feature subsets $P, Q \subseteq A$, *suppose* $n_P^\delta(x)$, $n_Q^\delta(x)$ *are two neighborhood granules of* x *on* P, Q *respectively, and* $n_D^0(x)$ *is an equivalent granule on the decision feature* D, *then we have*

1. *If* $P \subseteq Q$, *then* $I(n_P^\delta(x), n_D^0(x)) \leq I(n_Q^\delta(x), n_D^0(x))$;

2. *If* $0 \leq \gamma \leq \delta \leq 1$ *then* $I(n_P^\gamma(x), n_D^0(x)) \geq I(n_P^\delta(x), n_D^0(x))$.

Proof. 1. It is known from the Proposition 1, if $P \subseteq Q$, then $n_P^\delta(x) \supseteq n_Q^\delta(x)$. Therefore, one can achieve that $n_P^\delta(x) - n_D^0(x) \supseteq n_Q^\delta(x) - n_D^0(x)$, $U - (n_P^\delta(x) - n_D^0(x)) \subseteq U - (n_Q^\delta(x) - n_D^0(x))$. Then we can achieve that $\frac{|U - (n_P^\delta(x) - n_D^0(x))|}{|U|} \leq \frac{|U - (n_Q^\delta(x) - n_D^0(x))|}{|U|}$. According to the definition of granule inclusion degree, one can achieve that $I(n_P^\delta(x), n_D^0(x)) = \frac{|U - (n_P^\delta(x) - n_D^0(x))|}{|U|}$, $I(n_Q^\delta(x), n_D^0(x)) = \frac{|U - (n_Q^\delta(x) - n_D^0(x))|}{|U|}$. Therefore, $I(n_P^\delta(x), n_D^0(x)) \leq I(n_Q^\delta(x), n_D^0(x))$ is achieved.

2. Similarly, it is easily proved by the Proposition 1. If $0 \leq \gamma \leq \delta \leq 1$, then $I(n_P^\gamma(x), n_D^0(x)) \geq I(n_P^\delta(x), n_D^0(x))$ is achieved.

Definition 9. Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For any sample $x \in U$ and a condition feature subset $P \subseteq A$, R_P is the neighborhood relation derived from the feature subset P, and the variable precision neighborhood lower and upper approximation sets of the decision feature D with respect to the condition feature P are defined as follows:

$$R_{*P}(D)_\delta^\alpha = \{x \in U | I(n_P^\delta(x), n_D^0(x)) \geq \alpha\}, 0.5 \leq \alpha \leq 1,$$

$$R_P^*(D)_\delta^\beta = \{x \in U | I(n_P^\delta(x), n_D^0(x)) > \beta\}, 0 \leq \beta < 0.5,$$

The variable precision neighborhood lower approximation set of D with respect to P is also called the variable precision neighborhood positive region, which is expressed as

$$POS_P^\alpha(D)_\delta = R_{*P}(D)_\delta^\alpha.$$

The variable precision neighborhood boundary region of D with respect to P is defined as

$$BN_P^{\alpha,\beta}(D)_\delta = R_P^*(D)_\delta^\beta - R_{*P}(D)_\delta^\alpha.$$

The variable precision neighborhood negative region is defined as

$$NEG_P^{\alpha,\beta}(D)_\delta = U - R_P^*(D)_\delta^\beta.$$

Apparently for any α, β, $R_{*P}(D)_\delta^\alpha \subseteq R_P^*(D)_\delta^\beta$. The couple $< R_{*P}(D)_\delta^\alpha, R_P^*(D)_\delta^\beta >$ is called a variable precision neighborhood rough set. The positive region $POS_P^\alpha(D)_\delta$ reflects the classification accuracy of a classification system. The parameter α indicates a toleration of the classification accuracy, and the parameter δ is a threshold value of granulation. The variable precision boundary region embodies the uncertainty of classification, and the variable precision neighborhood negative region reflects the error rate of classification.

Definition 10. Let a classification system be $CS = (U, A, D)$ and $\delta \in [0, 1]$ be a neighborhood parameter, given any variable precision threshold α, for any feature subset $P \subseteq A$, the variable

precision neighborhood dependence of decision feature with respect to condition feature is defined as

$$\varphi_P^\alpha(D)_\delta = \frac{|POS_P^\alpha(D)_\delta|}{|U|}.$$

The variable precision neighborhood dependence is similar to the variable precision positive region, which reflects the variation of classification accuracy.

Proposition 3. *Let a classification system be* $CS = (U, A, D)$, $\delta \in [0, 1]$ *is a neighborhood parameter. For any two feature subsets* $P, Q \subseteq A$, *the variable precision neighborhood positive region satisfies the following conclusions:*

1. *If* $P \subseteq Q$, *then* $POS_P^\alpha(D)_\delta \subseteq POS_Q^\alpha(D)_\delta$;

2. *If* $0 \leq \gamma \leq \delta \leq 1$, *then* $POS_P^\alpha(D)_\gamma \supseteq POS_P^\alpha(D)_\delta$.

Proof. 1. From $P \subseteq Q$, according to the Proposition 2 we can know that $I(n_P^\delta(x), n_D^0(x)) \leq I(n_Q^\delta(x), n_D^0(x))$. By the definition of variable precision neighborhood lower approximation, we know $R_{*P}(D)_\delta^\alpha \subseteq R_{*Q}(D)_\delta^\alpha$. Therefore, $POS_P^\alpha(D)_\delta \subseteq POS_Q^\alpha(D)_\delta$ is achieved.

2. From $0 \leq \gamma \leq \delta \leq 1$, according to the Proposition 2 we can know that $I(n_P^\gamma(x), n_D^0(x)) \geq I(n_P^\delta(x), n_D^0(x))$. By the definition of variable precision neighborhood lower approximation, we know $R_{*P}(D)_\gamma^\alpha \supseteq R_{*P}(D)_\delta^\alpha$. Therefore, $POS_P^\alpha(D)_\gamma \supseteq POS_P^\alpha(D)_\delta$ is achieved.

According to the Proposition 3, we can draw the following conclusions:

Proposition 4. *If* $P_1 \subseteq P_2 \subseteq ... \subseteq P_m \subseteq A$, *then* $\varphi_{P_1}^\alpha(D)_\delta \leq \varphi_{P_2}^\alpha(D)_\delta \leq ... \leq \varphi_{P_m}^\alpha(D)_\delta \leq \varphi_A^\alpha(D)_\delta$.

Proposition 5. *If* $\delta_1 \leq \delta_2 \leq ... \leq \delta_m$, *then* $\varphi_P^\alpha(D)_{\delta_1} \geq \varphi_P^\alpha(D)_{\delta_2} \geq ... \geq \varphi_P^\alpha(D)_{\delta_m}$.

Proposition 6. *If* $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_m$, *then* $\varphi_P^{\alpha_1}(D)_\delta \leq \varphi_P^{\alpha_2}(D)_\delta \leq ... \leq \varphi_P^{\alpha_m}(D)_\delta$.

The Proposition 4 shows that the classification accuracy is increasing with the growing of features. It is not strictly monotonously increasing, so it may grow to a certain extent, the accuracy of classification no longer changing. Therefore, some features can be gradually selected so that the classification accuracy of these features has been increased to the same classification accuracy of all features. The Proposition 5 shows that the classification accuracy is decreasing with the increasing value of a granulation parameter. Because the process of granulation will increase noise, the larger the value of granulation parameter is, the more noise will be added and the accuracy of classification will decrease. The Proposition 6 shows that the accuracy of classification increases while the threshold of variable precision reduces. Therefore, it is necessary to consider reducing the threshold of the variable accuracy so as to improve the classification accuracy when the noise induced by the granulation parameter increases. The above analysis shows that the variable precision neighborhood rough set model is more robust to the classification systems and can handle the data containing noise.

4. FEATURE SUBSET SELECTION BASED ON VARIABLE PRECISION NEIGHBORHOOD ROUGH SETS

The information entropy, mutual information, rough entropy and knowledge granularity induced by equivalence relations measure the roughness of knowledge and reflect the classification ability of a decision system. However, these measures are suitable to deal with symbolic data. As for real-value data, we should find new methods. From the definition of variable precision neighborhood dependence, we put forward the concept of variable precision neighborhood reduction for real-value and noisy data in classification systems. Furthermore, a feature subset selection algorithm based on variable precision neighborhood rough sets is designed.

4.1. Feature Reduction Based on Variable Precision Neighborhood Rough Sets

Definition 11. Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For a condition feature $p \in P$, if $\varphi_{P-\{p\}}^{\alpha}(D)_{\delta} \neq \varphi_{P}^{\alpha}(D)_{\delta}$, then the feature is an independent feature in P. Otherwise, the feature is a redundancy feature in P.

A redundant feature provides useless or even spurious classification information. We delete the feature from a classification system, so a more compact and concise feature representation is appeared. It will reduce the dimension and complexity of a big data system.

Definition 12. Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For a condition feature $P \subseteq A$, it is a feature reduction if it satisfies the following conditions:

1. $\varphi_{P}^{\alpha}(D)_{\delta} = \varphi_{A}^{\alpha}(D)_{\delta}$;

2. $\forall p \in P, \varphi_{P-\{p\}}^{\alpha}(D)_{\delta} < \varphi_{P}^{\alpha}(D)_{\delta}$.

The first condition means that the selected features and the whole features have the same classification accuracy; the second condition ensures that the selected feature subset does not have redundant features.

From the above analysis, it is known that the variable precision neighborhood dependence reflects the classification ability of a feature subset. And according to the Proposition 4, the variable precision neighborhood dependence monotonically increases with the growing of the features. Therefore, the significance of a feature can be defined based on the variable precision neighborhood dependence.

Definition 13. Let a classification system be $CS = (U, A, D)$, $\delta \in [0, 1]$ is a neighborhood parameter. For a condition feature subset $P \subseteq A$ and a feature $p \in A - P$, the feature significance of p with respect to P is defined as

$$sgn_{\delta}^{\alpha}(p, P, D) = \varphi_{P \cup \{p\}}^{\alpha}(D)_{\delta} - \varphi_{P}^{\alpha}(D)_{\delta}.$$

The significance of a feature represents a change of the variable precision neighborhood dependence, before and after the feature is deleted. The threshold of variable precision shows the tolerance to the noise data.

4.2. Feature Subset Selection Algorithm Based on Variable Precision Neighborhood Rough Sets

From the Propositions 3 and 4, we know that the positive region and dependence of variable precision neighborhood are monotonic. Therefore, we can use the deleting or adding method to select the features. According to the definition of the variable precision neighborhood reduction, we propose a feature subset selection algorithm based on the variable precision neighborhood rough sets. Following the significance of feature as a heuristic information, a bottom-up feature subset selection algorithm is designed.

Algorithm 1: Algorithm VPNRS. A heuristic algorithm to feature selection based on variable precision neighborhood rough setsInput: A classification system $CS = (U, A, D)$, a neighborhood parameter $\delta \in [0, 1]$ and a variable precision threshold $0.5 \leq \alpha \leq 1$.

Output: A feature reduction RED.
(1): Initialization: $RED = \emptyset, T = \emptyset, \varphi_{RED}^{\alpha}(D)_{\delta} = 0, flag = true$
(2): while (*flag*)
(3): $T = RED, P = A - RED$
(4):   for each $p_k \in P$
(5): $T = RED \cup \{p_k\}, max = 0, POS = \emptyset$
(6):   for each $x_i \in U$
(7):   Computing the condition neighborhood granule
                 $n_T^{\delta}(x_i)$ by the Euclidean distance
(8):   Computing the decision granule $n_D^0(x_i)$
(9):   Computing the granule inclusion degree by
                 $$I(n_T^{\delta}(x_i), n_D^0(x_i)) = \frac{\left|U - (n_T^{\delta}(x_i) - n_D^0(x_i))\right|}{|U|}$$
(10):   If $I(n_T^{\delta}(x_i), n_D^0(x_i)) \geq \alpha$, then $POS = POS \cup \{x_i\}$
(11):   End for
(12): $POS_T^{\alpha}(D)_{\delta} = POS$
(13):   Computing the variable precision neighborhood dependence
                 $$\varphi_T^{\alpha}(D)_{\delta} = \frac{\left|POS_T^{\alpha}(D)_{\delta}\right|}{|U|}$$
(14):   If $\varphi_T^{\alpha}(D)_{\delta} > max$, then
(15): $max = \varphi_T^{\alpha}(D)_{\delta}, maxfeature = p_k$
(16):   End if
(17):   End for
(18): $p_k = maxfeature$
(19):   Computing $sgn_{\delta}^{\alpha}(p_k, RED, D) = \varphi_{RED \cup \{p_k\}}(D)_{\delta} - \varphi_{RED}^{\alpha}(D)_{\delta}$
(20):   If $sgn_{\delta}^{\alpha}(p_k, RED, D) \geq 0$, then
(21): $RED = RED \cup p_k$
(22): $\varphi_{RED}^{\alpha}(D)_{\delta} = \varphi_{RED \cup p_k}^{\alpha}(D)_{\delta}$
(23):   If $sgn_{\delta}^{\alpha}(p_k, RED, D) = 0$, then
(24): $flag = false$
(25):   End if
(26):   End if
(27): End while
(28): Return RED

In steps 6 and 7, achieving neighborhood granules costs $O(n^2 * m)$, where n is the number of samples and m is the number of features. Further, in step 4, it costs $O(m)$. Therefore, the time complexity of Algorithm VPNRS is $O(n^2 * m^2)$.

5. EXPERIMENTAL RESULTS

In this paper, some experiments are tested on four UCI datasets, which are shown in Table 1.

Due to the different value ranges of the datasets in the table, the datasets need to be normalized and preprocessed. We use the maximum-minimum method to ensure that all data are transformed into data between 0 and 1. The maximum–minimum normalization is defined as

$$f(x_i) = \frac{x_i - x_{min}}{x_{max} - x_{min}}.$$

Experimental results are compared from two aspects. One is the redundancy of selected feature subsets, and the other is the classification accuracy of reduced feature subsets. In the experiments, the values of neighborhood parameter start from 0.05 to 0.5, while their intervals are 0.05 and a total of 10 times are tested. In this paper, the algorithm needs to set the variable precision threshold, which is tested from 0.5 to 1 with an interval of 0.05, and there are 10 kinds of accuracy are changed. First, the feature subset selection (feature reduction) is performed using a variable precision neighborhood rough set model. And then the classification experiments are performed on the selected feature subsets. The experimental dataset is randomly divided into ten parts, of which nine parts are used for

training and the remaining one part for testing. The dataset is tested with the LIBSVM classifier developed by Professor Lin Chih-Jen.

5.1. Redundancy Comparison of Feature Reduction

The simplification of a feature reduction is measured by redundancy, and the redundancy is expressed as follows:

$$r = \frac{|RED|}{|A|}.$$

Among them, A is the set of all condition features, RED is a feature reduction and $|A|$ represents the cardinality of A. From this definition, we know that the smaller the redundancy, the better the simplification. The redundancy of a feature reduction varies with the changes of the neighborhood parameter and the variable precision threshold, and the experimental results are shown in Figure 2.

From the Figure 2(a), for the Glass dataset, when the neighborhood parameter δ sets 0.05–0.2 and the variable precision threshold α sets 0.55–0.75, the redundancy of feature reduction is smaller and the simplification effect is better. In contrast, the neighborhood parameter δ sets 0.35–0.5 and the variable precision threshold α sets 0.85–1, the redundancy of feature reduction is larger, and the simplification effect is poorer.
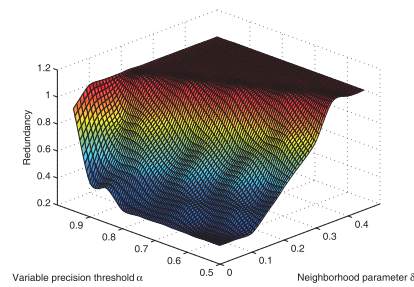
From the Figure 2(b), for the Pima dataset, when the neighborhood parameter δ sets 0.05–0.2 and the variable precision threshold α sets 0.55–0.75, the redundancy of feature reduction is smaller, and the simplification effect is better. In contrast, the neighborhood parameter δ sets 0.4–0.5 and the variable precision threshold α sets 0.8–1, the redundancy of feature reduction is larger, and the simplification effect is poorer.
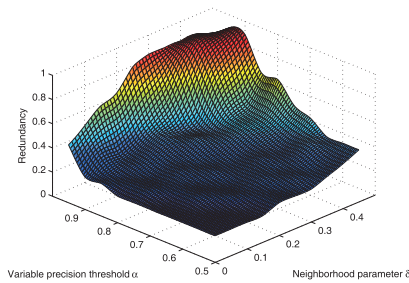
Table 1 | Experimental datasets.

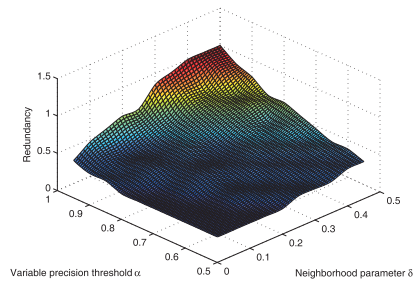| *Datasets*	$|U|$	$|A|$	$|V_D|$
Glass	214	9	6
Pima	768	8	2
Segmentation	210	19	7
Wine	178	13	3
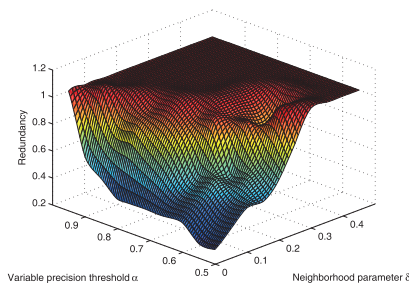
(a) Glass

(b) Pima

(c) Segmentation.

(d) Wine

Figure 2 | Redundancy comparisons of different datasets.

From the Figure 2(c), for the Segmentation dataset, the neighborhood parameter δ sets 0.05–0.35 and the variable precision threshold α sets 0.55–0.85, the redundancy of feature reduction is smaller, and the simplification effect is better. In contrast, the neighborhood parameter δ sets 0.4–0.5 and the variable precision threshold α sets 0.9–1, the redundancy of feature reduction is larger, and the simplification effect is poorer.

From the Figure 2(d), for the Wine dataset, when the neighborhood parameter δ sets 0.05–0.35 and the variable precision threshold α sets 0.55–0.9, the redundancy of feature reduction is smaller, and the simplification effect is better. In contrast, the neighborhood parameter δ sets 0.4–0.5 and the variable precision threshold α sets 0.95–1, the redundancy of feature reduction is larger, and the simplification effect is poorer.

From the above analysis, we can see that the smaller the neighborhood parameter is, then the smaller the reduction subset redundancy is and the better the reduction effect is. The smaller the variable precision threshold is, then the smaller the reduction subset redundancy is and the better the reduction effect is. Conversely, the larger the neighborhood parameter is, then the larger the reduction subset redundancy is and the worse the reduction effect is. The bigger the variable precision threshold is, then the larger the redundancy of the reduction subset is and the worse the reduction effect is. At the same time, this is consistent with the principle of Propositions 5 and 6. When the neighborhood parameter becomes bigger, the noise increases. According to the Proposition 5, the classification accuracy is reduced, and more features are needed to describe the dataset. When the variable precision threshold is smaller, according to Proposition 6, the classification accuracy increases, and fewer features can describe the dataset.

5.2. Comparison of Classification Accuracy

The smaller the redundancy, the less the number of features used for a classifier, which is beneficial to speed up the training of the classifier. However, the classification accuracy may be reduced and the effect of classification is not ideal when the selected features are less. Only when the selected feature subset satisfies the two conditions with small redundancy and high accuracy of classification, it is an ideal feature subset. Firstly, we get multiple feature subsets under different neighborhood parameter values and different thresholds, and then test the classification results of these feature subsets. The classifier is LIBSVM, and the experimental results are shown in Figure 3.

From the Figure 3(a), for the Glass dataset, when the neighborhood parameter sets 0.15 and the variable precision threshold sets 0.55, 0.6, and 0.7, the classification accuracy reaches the maximum value with 0.7143. When the neighborhood parameter sets 0.2 and the variable precision threshold sets 0.55, the maximum classification accuracy reaches 0.7143. When the neighborhood parameter sets 0.1 and the variable precision threshold sets 0.7, the classification accuracy reaches the minimum value with 0.4286.

From the Figure 3(b), for Pima datasets, the classification accuracy has 9 maximum values, mainly concentrated on a diagonal line of the plane formed by neighborhood parameters and variable precision thresholds. They are scattered in points with large neighborhood parameter values and small variable precision values, or with small neighborhood parameter values and large variable precision values. The minimum classification accuracy is distributed on the plane with small neighborhood parameter values and small variable precision values at the same time.

(a) Glass

(b) Pima

(c) Segmentation.

(d) Wine

Figure 3 | Classification accuracies of different datasets.

From the Figure 3(c), for Segmentation dataset, the classification accuracy has 3 maximum values, scattered in (δ, α) at (0.45,0.8), (0.45,0.85) and (0.4,1) points. The minimum classification accuracy is distributed on the plane with small neighborhood parameter values and small variable precision values at the same time.
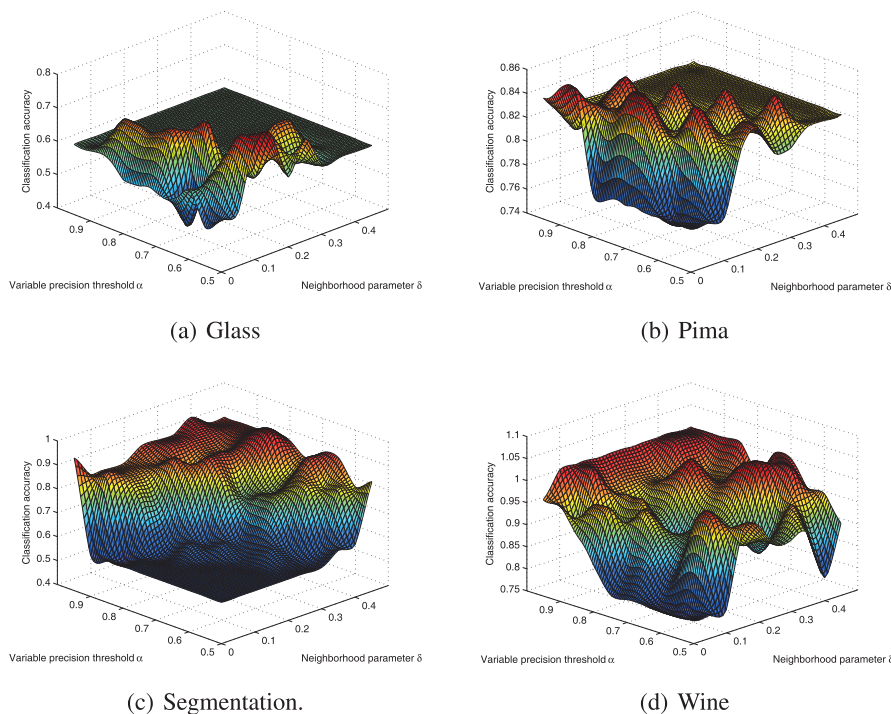
From the Figure 3(d), for Wine datasets, the points of maximum classification accuracy are numerous, mainly concentrated on a striped line with the variable precision value setting 1 and the neighborhood parameter ranging in 0.1–0.5. The points of minimum classification accuracy are distributed in a striped line with the small neighborhood parameter values and the variable precision threshold ranging in 0.55–0.8.

From the previous experiments, it is known that the simplification effects of feature subsets are better when the neighborhood parameter values and the variable precision thresholds are small. However, as shown in the Figure 3, the classification accuracies are smaller when the neighborhood parameter values and the variable precision thresholds are small. Therefore, the simplification of a feature subset is relevant to the performance of a classification. Meanwhile, from the Figure 3, we can see that when neighborhood parameter values and variable precision thresholds are large, there are many features for classification, but the accuracy of classification is not the best. Because that the neighborhood granulation will bring much noise to the datasets. When the neighborhood parameter value becomes larger, the noise is increased. Although the features are more, the noise is added due to the increasing of neighborhood parameter values, so the classification accuracy is affected. From the Proposition 6, we can see that the decreasing of the variable precision threshold can improve the classification accuracy. Since the noise introduced by a granulation process affects the classification accuracy, we can improve this situation by decreasing the threshold of variable precision, which counteracts the influence of the added noise. Therefore, it is necessary to adjust two values of neighborhood parameter and variable precision threshold at the same time, and select the appropriate features. It will be a good classification performance.

5.3. Comprehensive Comparisons of Redundancy and Classification Accuracy

The experiments in this subsection are compared from two aspects: redundancy of the selected features and their classification accuracies. In general, the classification accuracy is more important than redundancy. So we use the golden split point to set their weights. Suppose the redundancy of a feature subset is r, and its classification accuracy is c, we define a new comprehensive evaluation index $h = (1 - 0.618) * (1 - r) + 0.618 * c$, where $(1 - r)$ represents simplification. The experimental results are shown in Figure 4.

From the Figure 4(a), for the Glass dataset, there are 3 maximum values of the comprehensive index, which are distributed on the points (δ, α) of (0.15,0.55), (0.2,0.55) and (0.15,0.6). The minimum values of the comprehensive index are mainly distributed on the plane with δ ranging in 0.25–0.55 and α ranging in 0.8–1.

From the Figure 4(b), for the Pima dataset, there are 4 maximum values of the comprehensive index, which are distributed on the points (δ, α) of (0.05, 0.9), (0.05, 0.95), (0.1, 0.75) and (0.1, 0.8). The minimum values of the comprehensive index are mainly distributed on the plane with δ ranging in 0.3–0.55 and α ranging in 0.75–1.

From the Figure 4(c), for the Segmentation dataset, there is 1 maximum value of the comprehensive index, which is distributed on the point (δ, α) of (0.05, 1). There are 3 minimum values of the comprehensive index, which are distributed on the points (δ, α) of (0.4, 0.95), (0.45, 0.95) and (0.5, 0.95).

From the Figure 4(d), for the Wine dataset, there are 5 maximum values of the comprehensive index, which are distributed on the points (δ, α) of (0.2, 0.55), (0.25, 0.55), (0.15, 0.6), (0.2, 0.6) and (0.15, 0.65). There are 3 minimum values of the comprehensive index, which are distributed on the points (δ, α) of (0.4,1), (0.45,1) and (0.5,1).

In summary, the maximum values of comprehensive index are mainly distributed on the points of smaller neighborhood parameter values. As for the variable precision threshold, it is not very sensitive. The smaller the neighborhood parameter is, the smaller the noise introduced. Therefore, the comprehensive performance of classification accuracy and simplification is not very sensitive to the variable precision threshold under the condition of low noise. The minimum values of the comprehensive index are mainly distributed on the points where the neighborhood parameter values and the variable precision thresholds are both large. If the neighborhood parameter value is large, the noise introduced is much, which indicates that under the condition of much noise, the comprehensive performance of classification accuracy and simplification is more sensitive to the variable precision threshold. At this time, we need to decrease the threshold of variable precision, so that the performance of classification accuracy and simplification is improved.

In order to verify the advantage of variable precision neighborhood rough sets (VPNRS) in processing noisy data, we add Gaussian noise with (0,0,0.004), where the mean is 0, the variance is 0 and the noise amount is 0.004. We compare VPNRS with Hu's neighborhood rough set (NRS) [17] on these noisy data using K-nearest neighbor (KNN) and SVM classifiers. First, the NRS and VPNRS are utilized for feature selection, and then the KNN and SVM are used for classification experiments to compare their classification accuracies. The results are shown in Table 2.

As it can be seen from the Table 2, for the noisy data, the classification accuracies of VPNRS on both classifiers are higher than those of NRS, indicating that our proposed algorithm has obvious anti-noise ability.
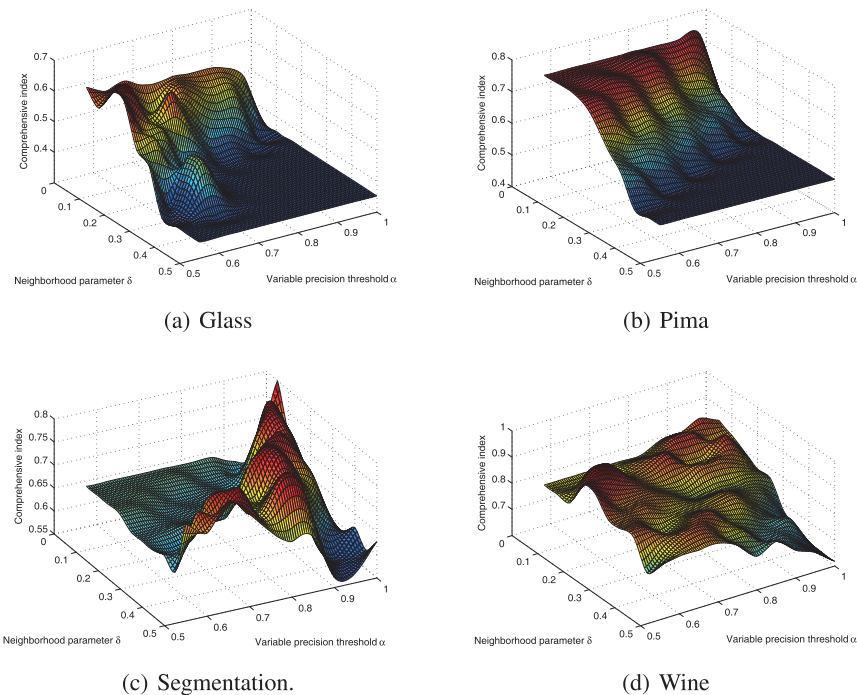
6. CONCLUSIONS

Aiming at the problem of traditional rough sets hardly dealing with real data, the neighborhood rough set granulation is introduced

Table 2 | Classification accuracies of UCI datasets with Gaussian noise.

Datasets	NRS + KNN	NRS + SVM	VPNRS + KNN	VPNRS + SVM
Glass	65.12±1.78	66.72±1.33	66.35±2.20	65.35±1.89
Pima	79.26±2.01	78.39±1.67	80.02±2.15	79.37±2.41
Segmentation	85.12±2.12	86.04±2.78	86.07±2.92	87.39±1.97
Wine	91.16±3.23	92.13±3.06	91.23±3.71	93.67±3.05

(a) Glass

(b) Pima

(c) Segmentation.

(d) Wine

Figure 4 | Comprehensive comparisons of different datasets.

to construct some condition neighborhood granules and decision equivalent granules in decision systems. The size of a granule is defined, the inclusion relation between a condition granule and a decision granule is analyzed, and a variable precision neighborhood rough set model based on the inclusive relation is proposed. Furthermore, the model is used to select feature subsets and a feature subset selection algorithm based on this model is designed. The experimental results show that the new proposed feature subset selection method can remove redundant features in decision systems. The method provides a simplified feature subset for the classifier and simultaneously maintains a good classification performance. It has certain robustness, and can handle the dataset containing noise. In the future work, we can apply the feature subset selection method proposed in this paper to the dimensionality reduction of big data systems.

CONFLICT OF INTEREST

We declare that we have no conflict of interest.

AUTHORS' CONTRIBUTIONS

Yingyue Chen: Conceptualization, methodology, software, validation, formal analysis, data curation, writing - original draft, project administration, funding acquisition, supervision, writing - review and editing, investigation, visualization, resources. Yumin Chen: Visualization, writing - review and editing.

FUNDING

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

[1] Z. Pawlak, Rough sets, Int. J. Inf. Comput. Sci. 11 (1982), 341–356.

[2] X. Jia, L. Shang, B. Zhou, *et al*., Generalized attribute reduct in rough set theory, Knowl. Based Syst. 91 (2016), 204–218.

[3] Y. Qian, J. Liang, W. Pedrycz, *et al*., Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (2010), 597–618.

[4] C.Z. Wang, Y. Wang, M.W. Shao, *et al*., Fuzzy rough attribute reduction for categorical data, IEEE Trans. Fuzzy Syst. 28 (2020), 818–830.

[5] W. Zhu, F. Wang, Reduction and axiomization of covering generalized rough sets, Inf. Sci. 152 (2003), 217–230.

[6] Y.M. Chen, D.Q. Miao, R.Z. Wang, A rough set approach to feature selection based on ant colony optimization, Pattern Recognit. Lett. 31 (2010), 226–233.

[7] Q. Hu, W. Pedrycz, D. Yu, *et al*., Selecting discrete and continuous features based on neighborhood decision error minimization, IEEE Trans. Syst. Man Cybern. Part B Cybern. 40 (2010), 137–150.

[8] F. Min, Q. Hu, W. Zhu, Feature selection with test cost constraint, Int. J. Approx. Reason. 55 (2014), 167–179.

[9] C.Z. Wang, Y. Huang, M.W. Shao, X.D. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl. Based Syst. 164 (2019), 205–212.

[10] J. Dai, Q. Xu, Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification, Appl. Soft Comput. 13 (2013), 211–221.

[11] P. Maji, S. Paul, Rough set based maximum relevance-maximum significance criterion and Gene selection from microarray data, Int. J. Approx. Reason. 52 (2011), 408–426.

[12] Y.F. Hassan, Rough sets for adapting wavelet neural networks as a new classifier system, Appl. Intell. 35 (2011), 260–268.

[13] D. Kim, Data classification based on tolerant rough set, Pattern Recognit. 34 (2001), 1613–1624.

[14] F. Pacheco, M. Cerrada, R.V. Snchez, *et al.*, Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery, Expert Syst. Appl. 71 (2017), 69–86.

[15] J. Zhan, Q. Wang, Certain types of soft coverings based rough sets with applications, Int. J. Mach. Learn. Cybern. 10 (2019), 1065–1076.

[16] Y.M. Chen, Y. Xue, Y. Ma, *et al.*, Measures of uncertainty for neighborhood rough sets, Knowl. Based Syst. 120 (2017), 226–235.

[17] Q. Hu, D. Yu, J. Liu, *et al.*, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (2008), 3577–3594.

[18] H. Zhao, P. Wang, Q. Hu, Cost-sensitive feature selection based on adaptive neighborhood granularity with multi-level confidence, Inf. Sci. 366 (2016), 134–149.

[19] Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inf. Sci. 111 (1998), 239–259.

[20] Q. Hu, D. Yu, Z. Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2008), 866–876.

[21] H.M. Chen, T.R. Li, Y. Cai, C. Luo, H. Fujita, Parallel attribute reduction in dominance-based neighborhood rough set, Inf. Sci. 373 (2016), 351–368.

[22] Y.M. Chen, Z.Q. Zeng, J.W. Lu, Neighborhood rough set reduction with fish swarm algorithm, Soft Comput. 21 (2017), 6907–6918.

[23] J. Li X. Yang X. Song, *et al.*, Neighborhood attribute reduction: a multi-criterion approach, Int. J. Mach. Learn. Cybern. 10 (2019), 731–742.

[24] Y. Liu, W.L. Huang, Y.L. Jiang, *et al.*, Quick attribute reduct algorithm for neighborhood rough set model, Inf. Sci. 271 (2014), 65–81.

[25] D.B. Chakraborty, S. Pal, Neighborhood rough filter and intuitionistic entropy in unsupervised tracking, IEEE Trans. Fuzzy Syst. 26 (2018), 2188–2200.

[26] Q. Hu, L. Li, P. Zhu, Exploring neighborhood structures with neighborhood rough sets in classification learning, in: A. Skowron, Z. Suraj (Eds.), Rough Sets and Intelligent Systems, Springer, Berlin, Heidelberg, Germany, 42 (2013), pp. 277–307.

[27] M.A. Tawhid, A.M. Ibrahim, Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm, Int. J. Mach. Learn. Cybern. 11 (2019), 573–602.

[28] F. Cao, J. Liang, G. Jiang, An initialization method for the K-means algorithm using neighborhood model, Comput. Math. Appl. 58 (2009), 474–483.

[29] P. Yao, Y.H. Lu, Neighborhood rough set and SVM based hybrid credit scoring classifier, Expert Syst. Appl. 38 (2011), 11300–11304.

[30] B.W. Zhang, F. Min, D. Ciucci, Representative-based classification through covering-based neighborhood rough sets, Appl. Intell. 43 (2015), 840–854.

[31] Y.M. Chen, Z.J. Zhang, J.Z. Zheng, *et al.*, Gene selection for tumor classification using neighborhood rough sets and entropy measures, J. Biomed. Informat. 67 (2017), 59–68.

[32] J. Meng, J. Zhang, Y.S. Luan, Gene selection integrated with biological knowledge for plant stress response using neighborhood system and rough set theory, IEEE/ACM Trans. Comput. Biol. Bioinformat. 12 (2015), 433–444.

[33] S.L. Wang, X. Li, S. Zhang, *et al.*, Tumor classification by combining PNN classifier ensemble with neighborhood rough set based gene reduction, Comput. Biol. Med. 40 (2010), 179–189.

[34] Y. Yu, W. Pedrycz, D.Q. Miao, Neighborhood rough sets based multi-label classification for automatic image annotation, Int. J. Approx. Reason. 54 (2013), 1373–1387.

[35] X.D. Yue, D.Q. Miao, N. Zhang, *et al.*, Multiscale roughness measure for color image segmentation, Inf. Sci. 216 (2012), 93–112.

[36] W. Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1993), 39–59.

[37] W. Ziarko, Probabilistic decision tables in the variable precision rough set model, Comput. Intell. 17 (2001), 593–603.

[38] T.H. Hou, C.C. Huang, Application of fuzzy logic and variable precision rough set approach in a remote monitoring manufacturing process for diagnosis rule induction, J. Intell. Manuf. 15 (2004), 395–408.

[39] I.T.R. Yanto, T. Herawan, M.M. Deris, Data clustering using variable precision rough set, Intell. Data Anal. 15 (2011), 465–482.

[40] D. Chen, Y. Yang, Z. Dong, An incremental algorithm for attribute reduction with variable precision rough sets, Appl. Soft Comput. 45 (2016), 129–149.

[41] M. Dash, H. Liu, Consistency-based search in feature selection, Artif. Intell. 151 (2003), 155–176.

[42] J. Golay, M. Leuenberger, M. Kanevski, Feature selection for regression problems based on the morisita estimator of intrinsic dimension, Pattern Recognit. 70 (2017), 126–138.

[43] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Trans. Fuzzy Syst. 17 (2009), 824–838.

[44] J. Li, H. Liu, Challenges of feature selection for big data analytics, IEEE Intell. Syst. 32 (2017), 9–15.

[45] M. Liang, X. Hu, Feature selection in supervised saliency prediction, IEEE Trans. Cybern. 45 (2017), 914–926.

[46] J. Dai, Q. Hu, H. Hu, *et al.*, Neighbor inconsistent pair selection for attribute reduction by rough set approach, IEEE Trans. Fuzzy Syst. 26 (2018), 937–950.

[47] Y. Li, S. Wu, Y. Lin, *et al.*, Different classes' ratio fuzzy rough set based robust feature selection, Knowl. Based Syst. 120 (2017), 74–86.

[48] Y. Yang, Incremental perspective for feature selection based on fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26(2017), 1257–1273.