WILEY | Hindawi

Research Article

Ensemble-Based Neighborhood Attribute Reduction: A Multigranularity View

Yuan Gao,[1] **Xiangjian Chen,**[1] **Xibei Yang** (ID)**,**[1] **Pingxin Wang,**[2] **and Jusheng Mi**[3]

[1]*School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China*
[2]*School of Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China*
[3]*College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang, Hebei 050024, China*

Correspondence should be addressed to Xibei Yang; jsjxy_yxb@just.edu.cn

Recently, multigranularity has been an interesting topic, since different levels of granularity can provide different information from the viewpoint of Granular Computing (GrC). However, established researches have focused less on investigating attribute reduction from multigranularity view. This paper proposes an algorithm based on the multigranularity view. To construct a framework of multigranularity attribute reduction, two main problems can be addressed as follows: (1) The multigranularity structure can be constructed firstly. In this paper, the multigranularity structure will be constructed based on the radii, as different information granularities can be induced by employing different radii. Therefore, the neighborhood-based multigranularity can be constructed. (2) The attribute reduction can be designed and realized from the viewpoint of multigranularity. Different from traditional process which computes reduct through employing a fixed granularity, our algorithm aims to obtain reduct from the viewpoint of multigranularity. To realize the new algorithm, two main processes are executed as follows: (1) Considering that different decision classes may require different key condition attributes, the ensemble selector is applied among different decision classes; (2) to accelerate the process of attribute reduction, only the finest and the coarsest granularities are employed. The experiments over 15 UCI data sets are conducted. Compared with the traditional single-granularity approach, the multi-granularity algorithm can not only generate reduct which can provide better classification accuracy, but also reduce the elapsed time. This study suggests new trends for considering both the classification accuracy and the time efficiency with respect to the reduct.

1. Introduction

Up to now, attribute reduction is one of the key topics in Rough Set Theory (RST) [1–4]. As a rough set-based feature selection, attribute reduction aims to reduce the data dimensions related to a given constraint. And the used constraint can be clearly explained by the considered measure such as approximation quality [5] and conditional entropy [6]. Actually, many established researches have employed different measures to design the corresponding algorithms. For example, Hu et al. [5] have studied uncertainty measures related to fuzzy rough set and then further explored approximation quality-based attribute reduction; Dai et al. [6–8] have investigated attribute reduction with respect to several types of conditional entropies. Therefore, the corresponding reducts are actually the subsets of the condition attributes which meet the constraints defined by using the measures such as approximation quality and conditional entropy, respectively.

These researches above have investigated different kinds of attribute reduction based on different measures, and these measures characterize the uncertainties or distinguishing abilities of the decision attribute through applying the information offered by the condition attributes. However, it is worth noting that they all focus on one and only one fixed parameter in attribute reduction [9–11], such as the fixed Gaussian kernel parameter in fuzzy rough set or the fixed radius in neighborhood rough set. From the viewpoint of Granular Computing (GrC) [12–16], applying one and only one parameter in rough set can only reflect the information

over a fixed granularity [17–19]. Therefore, the attribute reduction over a fixed granularity can be termed as the single-granularity attribute reduction. Nevertheless, the single-granularity attribute reduction fails to meet the requirements of practical applications. Generally speaking, the limitations of single-granularity attribute reduction can be mainly summarized as the following two aspects:

(1) The single-granularity attribute reduction fails to select attributes from multilevel [20–22] or multiview [23, 24]. For example, supposing that a reduct is obtained to preserve one given information granulation on the universe of discourse, this reduct may not be still the reduct for a little finer or coarser level of information granulation which may be caused by slight variation of data. In viewpoint of this, if it is requested to obtain a reduct with higher adaptability to the granularity world constructed by multilevel or multiview, it is necessary to establish the attribute reduction from the multigranularity view.

(2) The single-granularity attribute reduction may result in obtaining poor learning performance. The reduct derived from single-granularity algorithm may only acquire the ability from a fixed sight. It is mainly because the information over the fixed granularity may be lacking versatility and comprehensiveness. Actually, it is reported that given the same data sets, different granules can provide complementary predictive powers, and the learning performance may be improved by combining their information [25, 26]. To improve the learning performance, it may be a feasible method to employ the information provided by different granularities for computing a reduct. Therefore, it is necessary to generate a reduct through combining the information from the multigranularity view.

To overcome the limitations mentioned above, it is necessary to establish the attribute reduction from the multigranularity view. It contains two processes which are summarized as follows: (1) Construct the structure of multigranularity in RST. In this paper, the multigranularity structure will be constructed through employing different radii. As reported in [16, 27], the radius-based neighborhood forms a information granule, then the neighborhood-based singlegranularity can be constructed by applying one and only one radius, and it follows that neighborhood based multigranularity can be constructed by applying a set of different radii. In this process, applying a smaller radius will generate a finer information granule while applying a greater radius will generate a coarser information granule. The different scales of information granules can offer us the multigranularity based results of neighborhood rough approximations [25, 28]. Therefore, the structure of multigranularity is naturally obtained in neighborhood rough set. (2) Design a framework to realize the attribute reduction from the multigranularity view. The MultiGranularity View for Computing Reduct (MG-VCR) is constructed to realize the new algorithm, and the mechanism of MG-VCR is mainly faced up with two open problems which can be shown as follows:

(1) How to select candidate attribute: Generally speaking, all the decision classes are considered as a union in the traditional process of computing reducts, which can be seen as global view [29]. However, the traditional approach paid less attention to the key condition attributes related to different decision classes. In other words, the reduct derived from the traditional approach may lead to the improvement of one decision class but the deterioration of the other decision class. To fill such a gap, we will apply the ensemble selector [30] from local view. And, the new strategy pays much attention to the key condition attributes related to different decision classes [31]. Moreover, the majority voting strategy is employed in the ensemble selector among different decision classes for selecting candidate attribute.

(2) How to select granularity: Different from traditional single-granularity algorithms which focus on addressing the fixed problem over one given granularity, multigranularity attribute reduction is a framework which processes information from a new perspective. Actually, the multigranularity attribute reduction can consider the differences of the different granularities to generate the multigranularity reduct. Generally speaking, a natural way of finding multigranularity reduct is to generate reduct one by one in terms of each granularity. Nevertheless, employing all the granularities to generate reducts may be too time-consuming as the number of granularities increasing. To improve the time efficiency, given a set of granularities, fusing the information over some granularities of them may be an interesting attempt, such as only the finest granularity and the coarsest granularity are applied.

From the discussions above, the motivation of this paper is to construct a framework which can not only take the requirements of different decision classes into account but also realize the attribute reduction from multigranularity view. For one thing, to select attributes from the viewpoint of different decision classes, an ensemble selector [30] with majority voting strategy is applied. For another thing, to realize the attribute reduction from multigranularity view, a set of radii are employed to construct the neighborhood based multigranularity, then the new attribute reduction will be designed in terms of multigranularity view. The ensemble selector selects key condition attributes for different decision classes, and the multigranularity view may provide new strategies for more abilities related to the reduct. Immediately, the all process of constructing multigranularity attribute reduction is given in the following three steps:

(1) Firstly, the multigranularity structure is constructed in terms of the radius-based neighborhood. A set of radii sorted in ascending order are applied, and then different granularities can be obtained in turn, i.e., the neighborhood based multigranularity can be constructed.

(2) Secondly, to accelerate the process of computing reduct, the finest granularity and the coarsest

granularity instead of all the granularities are employed.

(3) Finally, the fitness function searching strategy is employed to compute reduct. In each iteration, the candidate attribute can be derived from all or most of the decision classes. In this process, the majority voting strategy is employed to determine which attribute to be added into the potential reduct.

The main contribution of this work can be summarized as the following aspects: (1) construct the neighborhood based multigranularity related to the given different radii; (2) design the attribute reduction from multigranularity view; (3) combine the ensemble strategy with the multigranularity attribute reduction; (4) the experimental results are conducted on 15 UCI data sets, and the experimental analyses demonstrate that our algorithms can be effective in the classification-oriented attribute reduction and efficient in the computational for generating reduct.

The structure of this paper is organized as follows. In Section 2, following some basic notions with respect to neighborhood rough set, the neighborhood based multigranularity is constructed. In Section 3, the frequently used measures and the traditional attribute reduction are presented. Section 4 introduces the ensemble selector from local view into the traditional framework firstly, and then the Multi-Granularity View for Computing Reduct (MG-VCR) is presented. The experimental results and the corresponding analyses are shown in Section 5. We conclude some remarks and perspectives for future research in Section 6.

2. Neighborhood-Based Multigranularity

In rough set theory, neighborhood [32] is an effective way to realize the information granule. It not only implies a simple way to express the information granules but also provides us a natural direction to construct the concept of the multigranularity.

Similar to other rough set theory [31, 33–36], the neighborhood rough set can also be formed in a decision system such that $\mathbb{D} = \langle U, AT \cup D \rangle$: U is the set of the samples, called the universe; AT is the set of the condition attributes which characterize properties of samples; D is the set of decision attributes for labeling samples. Note that the problem of single decision attribute is mainly discussed in this paper. Therefore, we use $D = \{d\}$ in which d is the considered single decision attribute. $\forall x \in U$ and $d(x)$ denotes the label or decision value of sample x. Assuming that the labels of samples are category, then an equivalence relation over d can be defined such that $\mathrm{IND}_d = \{(x, y) \in U \times U : d(x) = d(y)\}$, by IND_d, a set of the decision classes can be obtained such that $U/\mathrm{IND}_d = \{X_1, X_2, \ldots, X_q\}$. Furthermore, $\forall X_p \in U/\mathrm{IND}_d$ is referred to as the p-th decision class in rough set theory. Specially, the decision class which contains the sample x is denoted by $[x]_d$.

Definition 1. Given a decision system $\mathbb{D}$, $\forall A \subseteq AT$, consider a radius δ, then a neighborhood relation can be defined as

$$N_A^\delta = \{(x, y) \in U \times U : \Delta_A(x, y) \leq \delta\}, \quad (1)$$

where $\Delta_A(x, y)$ is the distance between samples x and y through using the information offered by A. Note that the Euclidean distance is employed in this paper.

Following equation (1), it is not difficult to construct the neighborhood of sample x such that

$$N_A^\delta(x) = \{y \in U : \Delta_A(x, y) \leq \delta\}. \quad (2)$$

Definition 2. Given a decision system $\mathbb{D}$, $U/\mathrm{IND}_d = \{X_1, X_2, \ldots, X_q\}$, $\forall A \subseteq AT$ and $\forall X_p \in U/\mathrm{IND}_d$, the neighborhood lower and upper approximations of X_p in terms of A are defined as

$$\underline{X_{P_A}}^\delta = \{x \in U : N_A^\delta(x) \subseteq X_p\};$$
$$\overline{X_{P_A}}^\delta = \{x \in U : N_A^\delta(x) \cap X_p \neq \varnothing\}. \quad (3)$$

From the viewpoint of the GrC, $N_A^\delta(x)$ can also be regarded as the neighborhood-based information granule [37–39]. Obviously, different values of δ can generate different results of $N_A^\delta(x)$. Supposing that a set of the radii $\mathbb{T} = (\delta^1, \delta^2, \ldots, \delta^n)$ in ascending order is considered, then given a sample $x \in U$, a set of the neighborhoods of x will be derived such that $(N_A^{\delta_1}(x), N_A^{\delta_2}(x), \ldots, N_A^{\delta_n}(x))$. These different scales of the neighborhoods imply a structure of multigranularity [25, 37, 40] (in this paper, a neighborhood and a granularity are equivalent terms). It is mainly because a smaller neighborhood expresses a finer granularity while a greater neighborhood expresses a coarser granularity. The details can be referred to Figure 1.

Figure 1 shows us the variation of the neighborhood with the radius increasing. It should be noticed that, in real-world applications, the samples in the given decision system may process n attributes, and then the corresponding neighborhood can be shown in n-dimensional situation. To simplify the description, Figure 1 shown is drawn in 2-dimensional situation. In sub-figure 1, given a radius δ_1, the samples x_1, x_2, x_3 are in the neighborhood of the testing sample y; in sub-figure 1, the radius is enlarged as δ_2, then more samples, i.e., samples x_4 and x_5, are included in the neighborhood of y; $\cdots$; finally, if the radius enlarged as δ_n in sub-figure 1, then more samples such that $x_6, x_7, x_8, x_9, x_{10}$ fall into the neighborhood of y. In this process, sub-figure 1 implies the finest granularity and sub-figure 1 implies the coarsest granularity.

As what has been pointed out in Figure 1, applying different radii can generate different granularities, then the multigranularity lower approximation and multigranularity upper approximation can be established through using different radii, the details expressions can be shown as

$$\underline{\mathbb{M}(X_P)}_A = \left(\underline{X_{P_A}}^{\delta_1}, \underline{X_{P_A}}^{\delta_2}, \ldots, \underline{X_{P_A}}^{\delta_n} \right);$$
$$\overline{\mathbb{M}(X_P)}_A = \left(\overline{X_{P_A}}^{\delta_1}, \overline{X_{P_A}}^{\delta_2}, \ldots, \overline{X_{P_A}}^{\delta_n} \right). \quad (4)$$
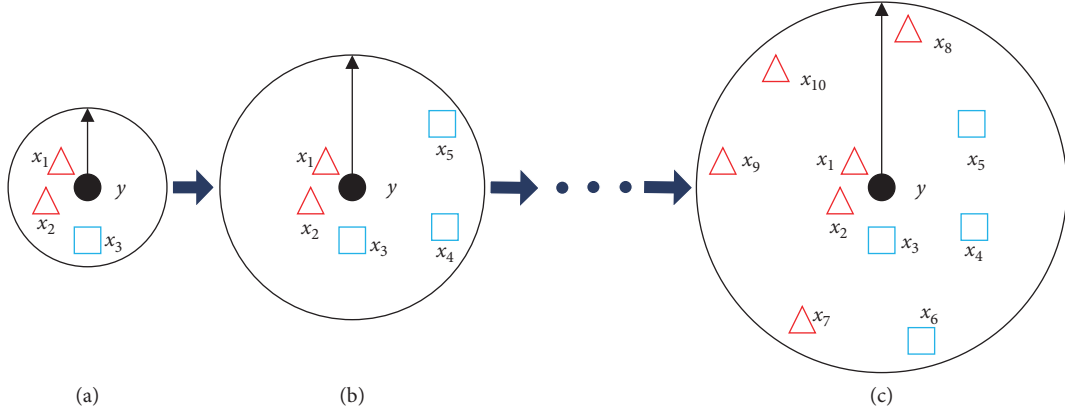
FIGURE 1: Variations of neighborhood.

3. Attribute Reduction

3.1. Some Measures. By considering the relation between neighborhood-based information granule and decision classes, a lot of measures have been explored in neighborhood rough set theory, especially for characterizing uncertainties and neighborhood-based attribute reduction. In this section, four measures will be addressed.

Approximation quality is one of the most frequently used measures, and it can be used to evaluate the certainty of belongingness by using the explanation of lower approximation in rough set model. The formal definition of approximation quality is shown as follows.

Definition 3. [41] Given a decision system $\mathbb{D}$, $U/\mathrm{IND}_d = \left\{ X_1, X_2, \ldots, X_q \right\}$, $\forall A \subseteq AT$, the approximation quality of d with respect to A is defined as

$$\gamma^\delta(A, d) = \frac{\left| \cup_{P=1}^{q} \underline{X_{P\_A}} \right|}{|U|}, \tag{5}$$

where $|\cdot|$ denotes the cardinality of the set.

Approximation quality reflects the percentage of the samples which belong to one of the decision classes by the explanation of lower approximations. Therefore, the greater the value of the approximation quality, the higher the degree of certain belongingness by the explanation of lower approximations in RST.

Conditional entropy is another widely accepted measure which can characterize the discriminating ability of condition attributes related to decision attribute. Presently, with respect to different requirements, various definitions of the condition entropies have been proposed [6–8, 42–44]. A typical form of the conditional entropy is shown as follows.

Definition 4. [44] Given a decision system $\mathbb{D}$, $\forall A \subseteq AT$, the conditional entropy of d with respect to A is defined as

$$E^\delta(A, d) = -\frac{1}{|U|} \sum_{x \in U} \left(\left| N_A^\delta(x) \cap [x]_d \right| \cdot \log \frac{\left| N_A^\delta(x) \cap [x]_d \right|}{\left| N_A^\delta(x) \right|} \right), \tag{6}$$

where the base of the logarithm "log" is 2.

The lower the value of the conditional entropy, the higher the discriminating ability of the condition attributes related to the decision attribute.

Similar to conditional entropy, neighborhood discrimination index [45] can also be used to characterize the ability of condition attributes to distinguish samples with different labels. However, conditional entropy requires the neighborhood based information granules derived from the neighborhood relation, and obtaining the neighborhood of each sample from neighborhood relation is time-consuming. To improve the time efficiency, Wang et al. [45] have proposed the measure of neighborhood discrimination index which can be directly obtained by the neighborhood relation instead of neighborhoods. And, the formal definition of the neighborhood discrimination index can be shown as follows.

Definition 5. [45] Given a decision system $\mathbb{D}$, $\forall A \subseteq AT$, the neighborhood discrimination index of d with respect to A is defined as

$$H^\delta(A, d) = \log \frac{\left| N_A^\delta \right|}{\left| N_A^\delta \cap \mathrm{IND}_d \right|}. \tag{7}$$

The lower the value of the neighborhood discrimination index, the higher the discriminating ability of the condition attributes related to the decision attribute.

Another measure deserved to be investigated is neighborhood decision error rate (NDER) which was proposed by Hu et al. [5]. NDER is commonly used to evaluate the classification performance of the neighborhood classifier (NEC) [46] in neighborhood rough set theory. Immediately, the formal definition of NDER can be defined as the follows.

Definition 6. [47] Given a decision system $\mathbb{D}$, $\forall A \subseteq AT$, the neighborhood decision error rate of d with respect to A is defined as

$$\mathrm{NDER}^\delta(A, d) = \frac{\left| \{ x \in U : \mathrm{Pre}(x) \neq d(x) \} \right|}{|U|}, \tag{8}$$

where $\mathrm{Pre}(x)$ is the predict label of the sample x derived from the used classifier.

The neighborhood decision error rate shows the percentage of samples which are misclassified with using neighborhood classifier.

3.2. Attribute Reduction.

Attribute reduction [48–53] is one of the key topics in rough set. Generally speaking, the purpose of attribute reduction is to remove the redundant attributes by a given constraint, and the constraint can be constructed by the measure such as approximation quality and conditional entropy. Up to now, many different types of the attribute reduction with respect to different requirements have been explored by researchers. However, it can be found that those definitions of attribute reduction may have similar structure. And, the general form of attribute reduction [54, 55] can be shown as follows.

Definition 7. Given a decision system $\mathbb{D}$, $\forall A \subseteq AT$, δ is a given radius, A is called the reduct in terms of the measure ψ if and only if

(1) A meets the ψ-constraint

(2) $\forall A' \subset A$, A' does not meet the ψ-constraint

In Definition 7, ψ is a given measure which can be "γ," "E," "H," or "NDER" in this paper. It is noted that the ψ-constraint should vary with the used measure ψ, and this can be further classified as two aspects:

(1) If a greater value is requested through using the measure ψ (such as $\psi = \gamma$), then the ψ-constraint can be set as no lower value obtained than using the raw attribute set, i.e., $\psi^{\delta}(A, d) \geq \psi^{\delta}(AT, d)$

(2) If a lower value is requested through using the measure ψ (such as $\psi = \mathrm{NDER}$), then the ψ-constraint can be set as no greater value obtained than using the raw attribute set, i.e., $\psi^{\delta}(A, d) \leq \psi^{\delta}(AT, d)$

Not only the definitions of attribute reduction have the similar structure but also the reducts can be derived by the similar searching strategy. Two main strategies are frequently used to compute reducts, i.e., exhaustive searching strategy and fitness function-based searching strategy. The former one is computationally more expensive than the latter one when dealing with data with the rapidly growing scale of samples, and the latter one is usually adopted due to its computational efficiency. It is mainly because different from the exhaustive strategy, fitness function-based searching strategy applies the greedy mechanism to compute reduct. Therefore, the reducts with respect to the measures shown in Section 3.1 will be computed by the similar fitness function-based searching strategy. To realize the fitness function-based searching strategy, the significances are required to evaluate the importance of the attributes. Then, the significance of attribute can be shown as follows:

$$\mathrm{Sig}_{\psi}^{\delta}(a_i, A, d) = \psi^{\delta}(A \cup \{a_i\}, d) - \psi^{\delta}(A, d); \tag{9}$$

$$\mathrm{Sig}_{\psi}^{\delta}(a_i, A, d) = \psi^{\delta}(A, d) - \psi^{\delta}(A \cup \{a_i\}, d), \tag{10}$$

where $\psi^{\delta}(A, d)$ is the value obtained by using the information over A in terms of d with using the measure ψ. The significance function shown in equatuon (9) indicates that the greater the value of the $\psi^{\delta}(A \cup \{a_i\}, d)$, the more important the attribute a_i is added into A; while the significance function shown in equation (10) indicates that the lower the value of the $\psi^{\delta}(A \cup \{a_i\}, d)$, the more important the attribute a_i is added into A.

As a frequently used method of this fitness function-based searching strategy, the addition method [30, 47] begins with an empty set and then selects the attribute with the maximum significance into the potential reduct in each iteration until the constraint is satisfied. Then, in the following context, we will use the addition method to further realize the fitness function-based searching strategy for computing reduct in terms of the measure ψ. Then, the addition method for computing reducts is shown in Algorithm 1.

Obviously, it is trivial to compute that the time complexity of Algorithm 1 is $O(|U|^2 \cdot |AT|^2)$. This can be attributed to the following two aspects: (1) the time complexity for computing neighborhood relation is $O(|U|^2)$ because any two samples in U should be applied for computing distance; (2) in the worst case, no attributes can be deleted from the raw attributes and then the iteration shown in Step 3 should be executed $|AT|$ times, and in the i-th iteration, Step 3 will be executed $|AT| - i + 1$ times. Finally, the time complexity of Algorithm 1 is $O(|U|^2 \cdot |AT|^2)$.

4. Ensemble-Based Attribute Reduction from Multigranularity View

4.1. Ensemble-Based Attribute Reduction.

Generally speaking, it is observed that the algorithms for computing reducts can be termed as a process of searching attributes. And, we can see from Algorithm 1 that when selecting attributes, the decision classes are considered as a union (global view), but the requirements of different decision classes have not been fully considered. In other words, more attention can be paid to the key condition attributes related to different decision classes (local view).

Presently, there is an increasing awareness of local view in rough set theory. This is mainly because local view may not only provide us more details of data but also offer us a direction to study the relationship among structures with respect to different decision classes. For instance, Chen and Zhao [29] have proposed the concept of local measures and investigated the structure of local attribute reduction in fuzzy rough set decision system, which provided a new insight into the problem of attribute reduction. Following this strategy, Song et al. [31] have investigated the local attribute reduction from the viewpoint of decision cost, different from the traditional approach which considered all decision classes from the global view, and they focused on the attributes with close connection to one given decision class than other ones. Yao and Zhang [53] have investigated the relationship between the classification-based attribute reduction (global view) and class-specific attribute reduction

---

Inputs: decision system $\mathbb{D}$, radius δ.
Outputs: a reduct: A.
(1) $A \longleftarrow \varnothing$;
(2) Compute $\psi^\delta(AT, d)$;
(3) **Do**
   (1) $\forall a_i \in AT - A$, compute $\mathrm{Sig}_\psi^\delta(a_i, A, d)$, select b such that $\mathrm{Sig}_\psi^\delta(b, A, d) = \max\{\mathrm{Sig}_\psi^\delta(a_i, A, d) : \forall a_i \in AT - A\}$;
   (2) $A = A \cup \{b\}$;
   Until A meets the given ψ-constraint;
(4) Return A.

ALGORITHM 1: Addition method for computing reduct (AMCR).

(local view). Though some previous results about local approaches have been explored in rough set theory, few of them have paid attention to the ensemble strategy from the local view. Therefore, to consider the different requirements of different decision classes, the ensemble strategy from local view may be a feasible solution. Immediately, the local measures [56] and the ensemble strategy from local view can be presented in the following context.

The flowing Definition 8 will show the local measures, and the four local measures corresponding to the measures shown in Section 3.1 can be defined as

Definition 8. Given a decision system $\mathbb{D}$, $U/\mathrm{IND}_d = \{X_1, X_2, \ldots, X_q\}$, $\forall A \subseteq AT$, $\forall X_p \in U/\mathrm{IND}_d$,

(1) The local approximation quality of X_p with respect to A is $\gamma^\delta(A, X_p) = |\underline{X_{P_A}}|/|X_p|$

(2) The local conditional entropy of X_p with respect to A is $E^\delta(A, X_p) = -(1/|X_p|)\sum_{x \in X_p}(|N_A^\delta(x) \cap [x]_d| \cdot \log(|N_A^\delta(x) \cap [x]_d|/|N_A^\delta(x)|))$

(3) The local neighborhood discrimination index of X_p with respect to A is $H^\delta(A, X_p) = \log(|N_A^\delta|/|N_A^\delta \cap X_p|)$

(4) The local neighborhood decision error rate of X_p with respect to A is $\mathrm{NDER}^\delta(A, X_p) = |\{x \in X_p : \mathrm{Pre}(x) \neq d(x)\}|/|X_p|$

Different from the measures shown in Section 3.1, the local measures only apply the specific decision class instead of all the decision classes together. For instance, the approximation quality shown in Section 3.1 reflects the percentage of the samples which belong to one of the decision classes by the lower approximations, while the local approximation quality reflects the percentage of the samples which belong to the decision class X_p by the corresponding lower approximation. Furthermore, the significances in terms of the decision class X_p can also be shown as follows:

$$\mathrm{Sig}_\psi^\delta(a_i, A, X_p) = \psi^\delta(A \cup \{a_i\}, X_p) - \psi^\delta(A, X_p); \quad (11)$$

$$\mathrm{Sig}_\psi^\delta(a_i, A, X_p) = \psi^\delta(A, X_p) - \psi^\delta(A \cup \{a_i\}, X_p). \quad (12)$$

The significance with respect to X_p shows the degree of importance of the attribute a_i in the individual decision class of X_p. Similarly, equations (11) and (12) are suitable for

different local measures; if a greater value is requested, then equation (11) is applied; otherwise, equation (12) is applied.

As the local measures and significances are presented above, then the ensemble strategy from the local view will be applied into the process of computing reduct. This process contains two key steps shown as follows:

(1) Each decision class selects an attribute with the maximal value of significance with respect to itself in each iteration

(2) The majority voting is employed to determine which attribute to be added into the potential reduct. It is noticed that the attribute with the highest frequency is selected as the candidate attribute each iteration, and if there are more than one attributes with the same highest frequency, the attribute which ranks lower/lowest will be selected

Take the reduct in terms of approximation quality as an example, the attribute with the maximal growth of the approximation quality is selected in each iteration using Algorithm 1; while in the new ensemble-based algorithm, the attribute with maximal growth of most of the lower approximations (local approximation quality) of different decision classes is selected. The reason comes from that if most of the lower approximations of decision classes have been increased, then the approximation quality in union can also increase. The following example will show the process of the ensemble strategy among different decision classes.

Example 1. As the decision system shown in Table 1, $U = \{x_1, x_2, \ldots, x_6\}$ is the set of samples, $AT = \{a_1, a_2, \ldots, a_6\}$ is the set of condition attributes, and d is the decision attribute.

It is obtained that $U/\mathrm{IND}_d = \{X_1, X_2, X_3\}$, and $X_1 = \{x_1, x_2\}$, $X_2 = \{x_3, x_4, x_5\}$, and $X_3 = \{x_6\}$.

Take the measure of "γ" as an example, supposing that $\delta = 0.08$, if AT is used, then it is obtained that $\gamma^\delta(AT, d) = 0.5000$.

If Algorithm 1 is applied, then the obtained reduct is $A_1 = \{a_4, a_5\}$ as $\gamma^\delta(A_1, d) = 0.6667 > \gamma^\delta(AT, d)$. Immediately, $\gamma^\delta(A_1, X_1) = 1.0000$, $\gamma^\delta(A_1, X_2) = 0.6667$, and $\gamma^\delta(A_1, X_3) = 0$ can be obtained. If the ensemble strategy is employed in the process for computing reducts, then the obtained reduct is $A_2 = \{a_1, a_2\}$. Actually, in the first iteration of this new process, the attribute with the maximum value of significance with respect to the decision class of

X_1, X_2, X_3 is a_1, a_5, a_2, respectively. This time the three attributes are with the same frequency. As a_1 ranks lowest in the first iteration, then a_1 is added into the potential reduct. In the next iteration, the selected attribute with respect to the decision class of X_1, X_2, X_3 is a_2, a_2, a_2, respectively, then a_2 is the attribute with the highest frequency. Therefore, a_2 is added into the potential reduct. Furthermore, the constraint is satisfied as $\gamma^\delta(A_2, d) = 0.8333 > \gamma^\delta(AT, d)$. Immediately, $\gamma^\delta(A_2, X_1) = 1.0000$, $\gamma^\delta(A_2, X_2) = 1.0000$, and $\gamma^\delta(A_2, X_3) = 0$ can be obtained.

It is observed that compared with the Algorithm 1, the ensemble strategy is employed, then the value of the local approximation quality with respect to the decision class X_2 has increasing, and the values of local approximation quality with respect to the decision class X_1 and X_3 remain unchanged. The above results tell us that the ensemble strategy among different decision classes can keep or improve the performance of single decision class. And, the similar mechanism can be applied for other measures.

4.2. Multigranularity Attribute Reduction.

Algorithm 1 shows the process of generating reducts over one and only one granularity, i.e., the algorithm from single-granularity view. As pointed in Section 1, the single-granularity attribute reduction has some inherent limitations. Therefore, in the following context, the attention can be paid to the multigranularity attribute reduction. And, the realization of multigranularity attribute reduction is made up with two key processes: (1) construct the multigranularity structure in RST; (2) realize the attribute reduction from the multi-granularity view. As shown in Section 2, given a set of radii $\mathbb{T} = (\delta_1, \delta_2, \ldots, \delta_n)$ which is in ascending order, it follows that different neighborhoods can be constructed by applying different radii. Furthermore, the multigranularity structure [28, 40, 57] can be constructed. Then, the multigranularity attribute reduction can be designed in the following contexts.

Definition 9. Given a decision system $\mathbb{D}$, $U/\text{IND}_d = \{X_1, X_2, \ldots, X_q\}$, $\forall A \subseteq AT$, an ascending order set $\mathbb{T}$, A is called the multigranularity reduct in terms of the measure ψ if and only if

(1) A meets the ψ^G-constraint

(2) $\forall A' \subset A$, A' does not meet the ψ^G-constraint

Similar to the Definition 7, the measure ψ can also be "γ," "E," "H," or "NDER," and the ψ^G-constraint is a multigranularity constraint which should also vary with the used measure ψ. But, the result should satisfy the multigranularity constraint instead of the single-granularity. Let us see the multigranularity constraint; a simple way to design the multigranularity constraint is to fuse all the constraints [25] with respect to all the considered granularities. However, it will bring us two challenges: (1) the complexity of the fused constraint will lower the speed of reduction process; (2) too many constraints will result in the difficulty of eliminating attributes. Therefore, to overcome the limitations above, we

TABLE 1: An example of decision system.

ID	a_1	a_2	a_3	a_4	a_5	a_6	d
1	0.1419	0.8147	0.1576	0.0975	0.6557	0.7577	1
2	0.0357	0.9058	0.9706	0.2785	0.5357	0.7431	1
3	0.3922	0.5469	0.9572	0.5469	0.8491	0.9157	2
4	0.6555	0.6553	0.4854	0.7922	0.9340	0.7922	2
5	0.3721	0.6354	0.5432	0.8344	0.6787	0.9595	2
6	0.7922	0.3429	0.9572	0.9595	0.5768	0.0561	3

will develop a quick process which is based on the fusion related to the finest and the coarsest granularities. It is mainly because that, for a given testing sample, applying the finest granularity, the fewest neighbors are obtained and applying the coarsest granularity, the most neighborhoods are obtained. And, this performance can also be clearly observed from Figure 1 in Section 2.

Section 4.1 shows the ensemble-based attribute reduction with majority voting strategy, the context in section analyzes the multigranularity view, and then the new proposed algorithm will combine the multigranularity view with the ensemble strategy together. The new algorithm will be termed as Multi-Granularity View for Computing Reduct (MG-VCR). Similar to Algorithm 1, MG-VCR compute reducts through using the addition method. However, there are two main differences summarized as follows:

(1) MG-VCR selects attributes from the viewpoints of different decision classes instead of the union, and the ensemble selector with the majority voting strategy is employed to select the candidate attribute

(2) The multigranularity view will be employed in the MG-VCR, and as only the finest granularity and the coarsest granularity are applied, then given a radii set $\mathbb{T} = (\delta_1, \delta_2, \ldots, \delta_n)$, only δ_1 and δ_n are employed in the process of computing reducts.

The first one gives a general framework about how to select candidate attributes, and the second one shows the strategy about how to select granularities in MG-VCR. Then, the detailed process using the addition method will be designed as follows:

In Algorithm 2, the time complexity of computing reduct is $O(q \cdot |U|^2 \cdot |AT|^2)$, in which q is the number of decision classes. Step 3 is the key in the process of computing reducts, what should be emphasized is the context shown as follows:

(1) The finest and the coarsest granularity each takes up the weight η_{δ_t} to realize the multigranularity view, and the condition $\eta_{\delta_1} + \eta_{\delta_n} = 1$ holds.

(2) (i) the process of the ensemble selector among the different decision classes $X_p (p = 1, 2, \ldots, q)$; (ii) the majority voting mechanism which is used for selecting the candidate attributes. It is noted that the "Psig" is the weighted average values related to the decision class "X_p." In this process, we first compute the weighted average values related to each decision class through using the finest granularity and the coarsest granularity, and then the ensemble-based

---

Inputs: decision system $\mathbb{D}$, $U/\text{IND}_d = \{X_1, X_2, \ldots, X_q\}$, a set of radii $\mathbb{T} = (\delta_1, \delta_2, \ldots, \delta_n)$.
Outputs: a multigranularity reduct: A.
(1) $A \longleftarrow \varnothing$;
(2)   Compute $\psi^{\delta_t}(AT, d)$, $(t = 1, n)$;
(3)   **Do**
      (1) **For** $p = 1$ to q
           $\forall a_i \in AT - A$, compute $\text{Sig}_{\psi^{\delta_t}}(a_i, A, X_p)$;
           $\forall a_i \in AT - A$, $\text{PSig}_{\psi}(a_i^P, A, X_p) = \sum \eta_{\psi^{\delta_t}} \cdot (\text{Sig}_{\psi^{\delta_t}}(a_i, A, X_p))$;
           $\text{PSig}_{\psi}(a_j^P, A, X_p) = \max\{\text{PSig}_{\psi}(a_i^P, A, X_p) : \forall a_i \in AT - A\}$;
           End
      (2) Record the frequencies of each attribute a_j^P and select the one with the highest frequency: b;
      (3) $A = A \cup \{b\}$;
      (4) Compute $\psi^{\delta_t}(A, d)$;
      Until A meets the ψ^G-constraint;
(4)   Return A.

ALGORITHM 2: Multi-Granularity View For Computing Reduct (MG-VCR).

strategy shown in Example 1 can be applied to select the suitable attribute.

(3) The multigranularity constraint (ψ^G-constraint) requires the fusion value of $\psi^{\delta_t}(A, d)$, $(t = 1, n)$ meet the constraint with respect to the fusion value of the $\psi^{\delta_t}(AT, d)$, $(t = 1, n)$, respectively. The fusion of the considered different granularities can be weighted mean value or harmonic mean value, etc. It is noticed that the weighted mean value is used in our experiment.

Finally, it is noticed that if only one granularity is used in this process, the algorithm will degenerate into single-granularity process, i.e., Single-granularity View for Computing Reduct (SG-VCR). Given the radii set $\mathbb{T} = (\delta_1, \delta_2, \ldots, \delta_n)$, SG-VCR should execute n times to obtain all the reducts, while MG-VCR only need execute one which may reduce the elapsed time as it applies the multigranularity view.

The following example shows us a general process of the Multi-Granularity View for Computing Reduct (MG-VCR). Suppose that $\delta_1 = 0.03, \delta_n = 0.30$, and $\eta_{\delta_1} = \eta_{\delta_n} = 1/2$ ($\eta_{0.03} = \eta_{0.30} = 1/2$).

Example 2. As the decision system shown in Table 1, take approximation quality as an example, as $\gamma^{\delta_1}(AT, d) = 0.6667, \gamma^{\delta_n}(AT, d) = 0.3333$ can be obtained, then the fusion value by the raw attributes (AT) is 0.5000.

It is noted that the significances of attributes are the also the fusion values (weighted average values). As $q = 3$ can be obtained, the significances of each attribute will be computed 3 times to obtain the importance degree related to each decision classes. Immediately, the most important attribute related to each decision classes will be obtained. In this process, the attribute with the maximum fusion value of significance with respect to the decision class of X_1, X_2, X_3 is a_1, a_2, a_1, respectively. Then, a_1 is added into the potential reduct (A) using the majority voting strategy. This employs the same strategy shown in Example 1.

Then, we should know if the potential reduct meets the ψ^G-constraint. Finally, as $\gamma^{\delta_1}(A, d) = 0.6667, \gamma^{\delta_n}(A, d) = 0.5000$ can be obtained, the fusion value by the A is 0.5834 which is larger than the fusion value by the raw attributes. Therefore, the reduct derived from the Algorithm 2 is $A = \{a_1\}$.

5. Experimental Analysis

In this section, to validate the effectiveness of MG-VCR proposed in this paper, 15 UCI data sets have been employed to conduct the experiment. The basic description of data information is shown in Table 2. All the experiments have been carried out on a personal computer with Windows 7, dual-core 1.50 GHz CPU, 8 GB memory. The programming language is MATLAB R2016a.

Fivefold cross-validation (5-CV) is used for evaluating the effectiveness of the proposed new algorithm. 5-CV divides all samples into 5 groups of the same size, and then four groups compose the training set to compute the reduct, while the one group composes the testing set to obtain the results for comparisons. The algorithm process repeats 5 times in turn, and the mean values are recorded. Besides, in the previous works [30, 58], it has been experimental proved that [0.1, 0.3] is an optimal candidate interval for the radius, where most of the classifiers can get good classification performance; to this end, 10 different values of δ such that $0.03, 0.06, \ldots, 0.30$ are selected. For SG-VCR, all the 10 radii will be employed to compute 10 reducts. As only the finest and the coarsest granularities are employed in the MG-VCR, the radii applied are 0.03 and 0.30, and the weight values are set as $\eta_{\delta_1} = \eta_{\delta_n} = 1/2$ ($\eta_{0.03} = \eta_{0.30} = 1/2$) in this paper; it should be emphasized that the MG-VCR only obtain a reduct using the given radii.

In the following context, "single-granularity algorithm" and "multigranularity algorithm" can be applied to denote the algorithm of Single-granularity View for Computing Reduct (SG-VCR) and Multi-Granularity View for Computing Reduct (MG-VCR), respectively. And, the results related to different reducts can be classified as two types. (1)

TABLE 2: Descriptions of data sets.

ID	Data sets	Samples	Attributes	Decision classes
1	Breast cancer Wisconsin (diagnostic)	569	30	2
2	Breast tissue	106	9	6
3	Climate model simulation crashes	540	20	2
4	Dermatology	366	34	6
5	Fertility	100	9	2
6	Forest type mapping	523	27	4
7	Gesture phase segmentation	9901	18	5
8	Ionosphere	351	34	2
9	Molecular biology (promoter gene sequences)	106	57	2
10	QSAR biodegradation	1055	41	2
11	Seeds	210	7	3
12	Statlog (German credit data)	1000	24	2
13	Vertebral column	310	6	2
14	Wine	178	13	3
15	Wine quality	6497	11	7

APP, ENT, DIS, and ERR denote the results related to the reduct derived from single-granularity algorithm in terms of approximation quality, conditional entropy, neighborhood discrimination index, and neighborhood decision error rate, respectively. And, these reducts can be collectively termed as single-granularity reducts. (2) MGAPP, MGENT, MGDIS, and MGERR denote the results related to the corresponding reduct derived from the multigranularity algorithm in terms of approximation quality, conditional entropy, neighborhood discrimination index, and neighborhood decision error rate, respectively. And, these reducts can be collectively termed as multigranularity reducts.

5.1. Comparisons among the Lengths of Reducts.
Table 3 shows the comparisons among the lengths of reducts, and the reducts are derived from single-granularity algorithm and multigranularity algorithm in terms of different measures.

With a careful investigation of Table 3, it is not difficult to observe the following.

In the aspect of the lengths, the reducts derived from multigranularity algorithm are greater than those derived from single-granularity algorithm generally. In other words, more attributes are added into the reduct set with using the multigranularity algorithm. Though there may exist some fluctuations between the measures of "DIS" and "MGDIS," the differences between the lengths with respect to the "DIS" and "MGDIS" are not obvious. Furthermore, the average values in the last row can demonstrate that reducts derived from multigranularity algorithm need more attributes to be added into the reduct than single-granularity algorithm. To sum up, the multigranularity algorithm can almost coincide with that of single-granularity algorithm in the lengths of reducts.

5.2. Comparisons among the Elapsed Time.
Table 4 shows the comparisons among the elapsed time related to single-granularity algorithm and multigranularity algorithm in terms of different measures. It should be noticed that the elapsed time of the single-granularity algorithm is the summate of the processes through using the constraints related to the 10 granularities, and the elapsed time of the multigranularity algorithm is the one process using the fusion constraint related to the finest granularity and the coarsest granularity. In this experiment, the obtained multigranularity reduct will represent all the reducts related to all the given granularities, and the finest granularity and the coarsest granularity are only employed to select the attributes. Therefore, the multigranularity reduct is more than the reduct related to the used granularities, and it will be employed over all the given 10 granularities. Besides, the unit of the elapsed time is "s."

With a careful investigation of Table 4, it is not difficult to observe the following:

(1) The multigranularity algorithm requires less time compared with the corresponding single-granularity algorithm. Furthermore, the total elapsed time of single-granularity algorithm in terms of all the fixed radii is five times as much as the elapsed time of multigranularity algorithm generally. It is mainly because the single-granularity algorithm applied all the given granularitities for computing reducts while the multigranularity algorithm only applied the finest granularity and the coarsest granularity.

(2) There also exists an exception: in the data set "Wine Quality" (ID: 15), the elapsed time of computing "MGDIS" is greater than computing "DIS." This can mainly attribute to the too great value of the length; this can be observed in Table 3; the "MGDIS" is almost eight times as long as the "DIS."

(3) Besides, the average elapsed time listed in the last row demonstrates that the multigranularity algorithm requires less time than single-granularity algorithm. To sum up, the elapsed time in Table 4 demonstrates that the multigranularity algorithm can improve the time efficiency compared with single-granularity algorithm.

5.3. Comparisons among the Classification Accuracies Using NEC.
Figure 2 shows us the comparisons among classification accuracies with respect to the single-granularity reducts and the multigranularity reducts using NEC. It is noted that the multigranularity algorithm obtains one multigranularity reduct, and this reduct will represent all the reducts related to n ($n = 10$ in our experiment) granularities. Therefore, the one multigranularity reduct will be employed under all the granularities. In the following charts, the X-axis represents different kinds of reducts, and the Y-axis represents the corresponding values of classification accuracy.

With a careful investigation of Figure 2, it is not difficult to observe the following:

TABLE 3: Comparisons among the lengths of reducts (the greater values are in bold).

ID	APP	MGAPP	ENT	MGENT	DIS	MGDIS	ERR	MGERR
1	**4.5000**	4.0000	2.4400	**3.2000**	**30.0000**	25.0000	**21.5200**	18.6000
2	2.4800	**5.0000**	1.8600	**2.2000**	1.0200	**4.2000**	2.8400	**5.2000**
3	4.4400	**5.0000**	2.5600	**3.0000**	**20.0000**	9.6000	**1.5800**	1.0000
4	13.4200	**18.2000**	7.0400	**11.8000**	**32.1600**	26.8000	21.9600	**32.8000**
5	7.0200	**7.4000**	7.3800	**8.0000**	**9.0000**	8.4000	**1.1000**	1.0000
6	11.3400	**18.0000**	3.8600	**11.0000**	13.6400	**19.2000**	**9.0000**	7.8000
7	3.1600	**7.8000**	3.7800	**6.2000**	**18.0000**	16.2000	5.0400	**8.6000**
8	**9.5600**	7.0000	1.4200	**3.0000**	**12.8600**	6.2000	2.4200	**3.8000**
9	2.8800	**4.4000**	2.0000	**2.6000**	**7.0000**	**7.0000**	2.0000	**3.0000**
10	10.3600	**18.2000**	4.9200	**10.8000**	16.8200	**34.4000**	17.3200	**21.2000**
11	3.1000	**3.8000**	3.3400	**4.0000**	**7.0000**	6.2000	**7.0000**	4.0000
12	**11.0600**	10.8000	10.5000	**11.0000**	**24.0000**	12.6000	**3.1000**	2.2000
13	**4.0800**	3.6000	2.2000	**4.2000**	5.5000	**6.0000**	3.5600	**4.2000**
14	**6.4780**	5.2000	4.9200	**6.0000**	**13.0000**	12.2000	10.7400	**11.0000**
15	6.2400	**7.6000**	4.1800	**6.6000**	1.0000	**8.2000**	**1.0000**	1.0000
Avg	6.6745	**8.4000**	4.1600	**6.2400**	**14.0600**	13.4800	7.3453	**8.3600**

TABLE 4: Comparisons among the elapsed time.

ID	APP	MGAPP	ENT	MGENT	DIS	MGDIS	ERR	MGERR
1	71.4247	*13.7473*	66.0414	*11.1967*	301.0050	*50.9976*	220.0193	*39.6363*
2	0.6136	*0.2186*	1.1509	*0.1183*	0.2982	*0.1915*	0.7021	*0.2427*
3	32.0207	*7.0614*	33.5664	*4.7105*	88.3461	*11.0420*	11.6320	*1.5377*
4	127.4708	*31.6491*	124.9537	*22.5923*	209.1630	*38.6443*	173.5181	*41.5509*
5	0.5358	*0.1365*	1.2395	*0.1694*	0.6159	*0.1107*	0.1189	*0.0436*
6	213.2737	*60.5530*	139.8738	*45.7522*	196.7573	*60.2659*	191.1344	*35.8713*
7	$3.9206E+03$	*$1.8569E+03$*	$1.1352E+04$	*$1.4564E+03$*	$1.5761E+04$	*$2.8735E+03$*	$4.6174E+04$	*$1.3484E+04$*
8	43.4906	*7.0300*	14.2405	*3.1897*	99.5249	*6.4197*	12.1248	*3.8759*
9	0.7876	*0.2522*	2.5775	*0.1462*	10.2326	*2.0908*	1.5736	*0.4877*
10	280.6782	*54.6040*	341.0450	*72.6792*	118.3220	*31.5351*	675.3199	*150.4289*
11	2.7859	*0.4199*	3.1873	*0.4339*	3.1362	*0.5318*	2.1888	*0.4149*
12	358.8993	*75.2890*	513.9914	*69.0625*	537.2714	*75.0350*	110.9560	*15.7937*
13	2.7859	*0.5061*	3.1873	*0.5695*	3.1362	*0.6182*	2.1888	*0.5042*
14	4.8653	*0.9244*	7.3322	*0.9314*	6.6377	*1.3385*	5.8532	*1.2405*
15	$2.3221E+04$	*$1.6156E+04$*	$5.0364E+04$	*$5.7845E+03$*	$3.5816E+03$	*$1.7434E+04$*	$3.5871E+03$	*673.4737*
Avg	1885.4155	*1217.6861*	4197.8925	*498.1635*	1394.4698	*1372.4214*	3411.2287	*963.2735*

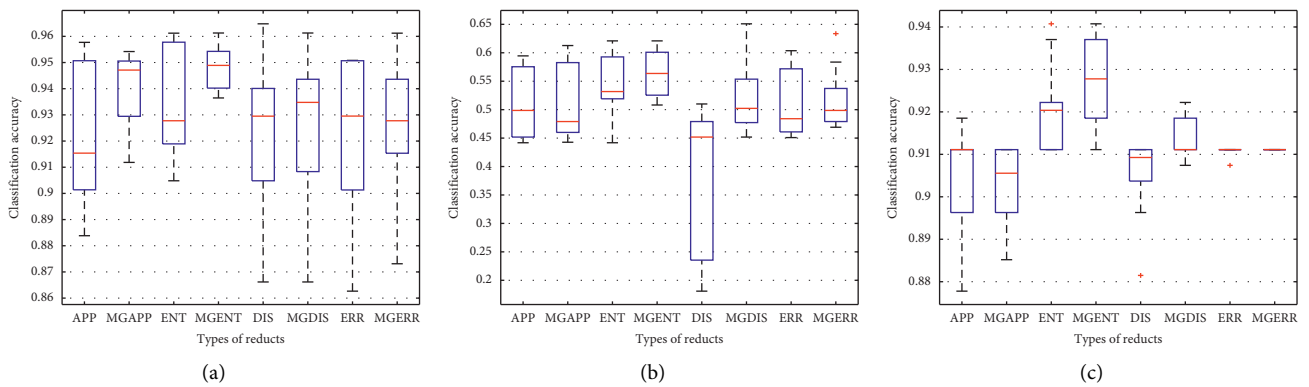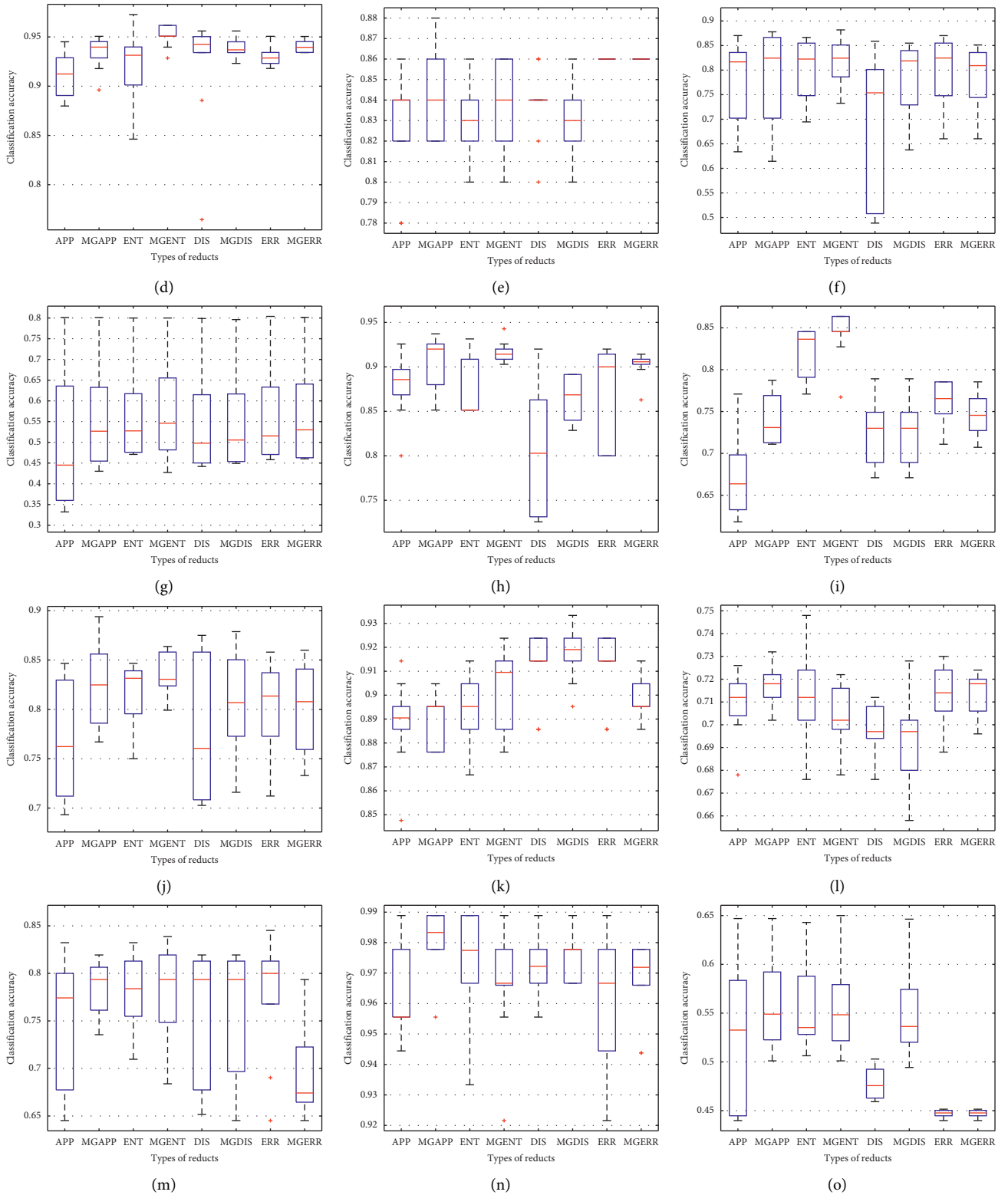The lower values are given in Italics.

FIGURE 2: Continued.

Figure 2: Comparisons among the classification accuracies with using NEC. (a) Breast Cancer Wisconsin (Diagnostic). (b) Breast Tissue. (c) Climate Model Simulation Crashes. (d) Dermatology. (e) Fertility. (f) Forest Type Mapping. (g) Gesture Phase Segmentation. (h) Ionosphere. (i) Molecular Biology (Promoter Gene Sequences). (j) QSAR Biodegradation. (k) Seeds. (l) Statlog (German Credit Data). (m) Vertebral Column. (n) Wine. (o) Wine Quality.

(1) The multigranularity reducts can improve classification accuracy compared with the single-granularity reducts. Take the data set of "Ionosphere" (ID: 8) as an example, the mean values of the classification accuracies derived from the single-granularity reducts are 0.8800, 0.8726, 0.8074, are 0.8686 with using APP, ENT, DIS, and ERR, respectively. While the mean values of the classification accuracies derived from the multigranularity reducts are 0.9057, 0.9154, 0.8646, and 0.9023 using MGAPP, MGENT, MGDIS, and MGERR, respectively. This demonstrates that the attribute reduction from multigranularity view can improve the classification accuracy.

(2) In the data set "Wine Quality" (ID: 15), the length of reduct shown in Table 3 indicates that more attributes are added into the reduct set using the reduct "DIS," while the mean value of classification accuracy in terms of "MGDIS" is also much greater than the value in terms of "DIS." Besides, the multigranularity reducts can generally improve the classification performance than the single-granularity reducts. This may indicate that though there exists some difference among different granularities, considering the finest and the coarsest granularities may be a feasible way to introduce the multigranularity thing into the attribute reduction.

Table 5 shows the comparisons among the classification accuracies related to the raw attributes, single-granularity reducts, and multigranularity reducts. To felicitate the readers, the greatest values of classification accuracy are in bold, and the smallest values are in italics.

With a careful investigation of Table 5, it is not difficult to observe the following:

(1) As for the measure related to the approximation quality and neighborhood discrimination index, among the comparisons, the largest value tends to be the one related to the multigranularity reduct, and the smallest value tends to be the one related to the single-granularity reduct. It implies that overfitting [59] may occur in most data sets using these single-granularity reducts. In other words, the single-granularity reducts perform better than the raw attributes in training samples but perform worse in testing samples.

(2) As for the measure related to conditional entropy and neighborhood decision error rate, among the comparisons, the largest value tends to be the one related to the multigranularity reduct, and the smallest value tends to be the one related to the raw attributes.

(3) To sum up, whether the singe-granularity reducts perform well or not, the multigranularity reducts can improve the classification performance.

The results shown in this subsection demonstrates that the multigranularity view may suggest new trends for improving the classification performances related to the reducts.

Furthermore, the Wilcoxon Rank Sum Test method will be employed for comparing the distributions with respect to the classification accuracies derived from the single-granularity algorithm and multigranularity algorithm. The purpose of our comparison is trying to reject the null-hypothesis that the distributions of the classification accuracies are significantly different. Assuming that the threshold is set as 0.05, if the value is greater than 0.05, we then reject the null-hypothesis, i.e., the distributions of the classification accuracies are similar.

It is observed from Table 6 that the obtained values are greater than 0.05 generally. As for the obtained values which are less than 0.05, we can find from Figure 2 that the mean values with respect to multigranularity reducts are greater than those with respect to the single-granularity reducts generally. Take the data set of "Wine Quality" (ID: 15) as an example, in the comparison of "DIS & MGDIS," the result with the method is less than 0.05, but the mean value of classification accuracy in terms of "DIS" and "MGDIS" is 0.4779 and 0.5507, respectively. The only exception is the data set "Vertebral Column" (ID: 13), in the comparison of "ERR & MGERR," the mean value of classification accuracy in terms of "ERR" and "MGERR" is 0.7768 and 0.6935, respectively. To sum up, there is not much significant difference between these two distributions of classification accuracies, and the multigranularity reducts even improve the classification performance than the single-granularity reducts.

5.4. Comparisons among the Classification Accuracies Using SVM and CART. To further verify the effectiveness of multigranularity algorithm, the other two popular classifiers are also applied which are SVM [60–62] and CART [63]. To make the comparison feasible, the mean classification accuracies with respect to the single-granularity reduct are computed by those responding 10 values.

Tables 7 and 8 display the comparisons among the classification accuracies with respect to raw attributes, the single-granularity reducts, and the multigranularity reducts using SVM and CART, respectively. To felicitate the readers, the greatest values of classification accuracy are in bold, and the smallest values are in italics.

With a careful investigation of Tables 7 and 8, it is not difficult to observe the following:

(1) In Table 7, the classification accuracies related to the raw attributes are largest generally. And, in Table 8, the percentage that the classification accuracies related to the raw attributes is largest and almost 55%. It implies that overfitting occurs in most data sets with using SVM or CART.

(2) The classification accuracies with respect to the multigranularity reducts are greater than the single-granularity reducts generally. Actually, in Table 7, among the comparisons (15 UCI data sets), the performance of "MGENT" is the best, the larger

Table 5: Comparisons among the classification accuracies with using NEC.

ID	RAW	APP	MGAPP	RAW	ENT	MGENT	RAW	DIS	MGDIS	RAW	ERR	MGERR
1	0.9232	*0.9204*	**0.9401**	*0.9232*	0.9337	**0.9489**	0.9232	*0.9232*	**0.9253**	0.9232	*0.9211*	**0.9253**
2	*0.5038*	0.5104	**0.5125**	*0.5038*	0.5389	**0.5600**	0.5038	*0.3748*	**0.5204**	*0.5038*	0.5131	**0.5176**
3	0.9052	0.9037	*0.9019*	0.9052	0.9211	**0.9274**	0.9052	0.9052	**0.9133**	0.9052	0.9107	**0.9111**
4	**0.9557**	0.9181	0.9373	0.9557	*0.9339*	**0.9561**	0.9557	*0.9406*	0.9470	0.9557	*0.9432*	**0.9575**
5	0.8380	0.8300	**0.8440**	0.8380	*0.8280*	**0.8400**	0.8380	0.8380	*0.8320*	0.8380	**0.8600**	**0.8600**
6	**0.7871**	0.7794	0.7843	*0.7871*	0.8030	**0.8156**	0.7871	*0.6843*	0.7821	**0.7871**	0.7965	*0.7843*
7	0.5454	*0.4987*	**0.5580**	*0.5454*	0.5634	**0.5743**	*0.5454*	0.5454	**0.5497**	*0.5454*	0.5602	**0.5678**
8	*0.8057*	0.8800	**0.9057**	0.8057	0.8726	**0.9154**	0.8057	0.8074	**0.8646**	0.8057	0.8686	**0.9023**
9	0.7235	0.6729	**0.7402**	0.7235	0.8235	**0.8431**	0.7235	0.7235	0.7235	*0.7235*	**0.7644**	0.7500
10	0.8089	0.7688	**0.8241**	0.8089	0.8158	**0.8339**	0.8089	*0.7762*	0.8068	**0.8089**	*0.8017*	*0.8017*
11	**0.9124**	0.8886	0.8905	0.9124	*0.8943*	**0.9038**	*0.9124*	*0.9124*	**0.9171**	0.9124	0.9124	*0.8990*
12	0.6986	0.7090	**0.7172**	0.6986	0.7120	**0.7038**	0.6986	0.6985	*0.6926*	0.6986	0.7134	**0.7144**
13	0.7639	0.7510	**0.7871**	0.7639	0.7794	**0.7800**	0.7639	0.7619	0.7639	0.7639	**0.7768**	*0.6935*
14	0.9733	0.9631	**0.9811**	0.9733	*0.9722*	**0.9656**	*0.9733*	0.9733	**0.9756**	0.9733	*0.9610*	0.9675
15	0.5500	0.5246	**0.5596**	0.5500	0.5500	**0.5574**	0.5500	*0.4779*	**0.5507**	0.5500	*0.4469*	*0.4469*
AVG	*0.7597*	0.7679	**0.7922**	*0.7597*	0.7961	**0.8084**	0.7597	*0.7562*	**0.7843**	0.7597	*0.7029*	**0.7799**

The greatest values are in bold, and the smallest values are in italics.

Table 6: Comparisons of the distributions with NEC.

ID	APP & MGAPP	ENT & MGENT	DIS & MGDIS	ERR & MGERR
1	0.1986	0.1503	0.7911	0.9397
2	0.5952	0.4717	0.1116	0.7617
3	0.3944	0.2516	*0.0430*	0.3681
4	*0.0153*	*0.0481*	0.7321	*0.0252*
5	0.0934	0.1115	0.9668	0.3681
6	0.7054	0.5851	0.1212	0.5706
7	0.3256	0.8205	0.6232	0.8501
8	0.0814	*0.0298*	0.0954	0.8474
9	*0.0045*	0.0510	1.0000	0.2043
10	0.2718	*0.0339*	0.6500	0.4707
11	1.0000	0.2685	0.5593	0.0636
12	0.1291	0.2558	0.6494	0.9396
13	0.2395	0.9095	0.9694	*0.0071*
14	*0.0098*	0.3777	0.7793	0.8780
15	0.4055	0.7913	*3.2984E − 04*	1.0000
Avg	0.2834	0.3645	0.5166	0.5862

The values less than 0.05 are italics.

values take up 80% (12 data sets); and as for "MGAPP," "MGDIS," and "MGERR," they all get 66.7% (10 data sets). The similar performance can be seen in Table 8, the performance of "MGDIS" and the "MGERR" is the best, and the larger values both take up 86.7% (13 data sets); and the suboptimal one is "MGAPP" which gets 73.3% (11 data sets), the performance of "MGENT" which gets 66.7% (10 data sets).

(3) To sum up, the multigranularity reducts are superior to those single-granularity reducts in improving classification accuracies. As presented in Tables 7 and 8, the multigranularity reducts can improve the classification accuracies generally. And the average values of the 15 UCI data sets in both the tables can also demonstrate that the classification accuracies related to the multigranularity reducts are greater than the single-granularity reducts. To sum up, the multigranularity reducts perform better than those single-granularity reducts with using SVM and CART.

The Anoval function was used to verify whether there is a significant difference existing in the distributions with respect to the classification accuracies. Similar to the Wilcoxon Rank Sum Test shown above, the purpose of the comparison is trying to reject the null-hypothesis that the distributions of the classification accuracies are significantly different. Assuming that the threshold is set as 0.05 and if the value is greater than 0.05, we then reject the null-hypothesis, i.e., the distributions of the classification accuracies are similar. And the results of the comparison among the distributions using SVM and CART are listed in the following Tables 9 and 10, respectively.

Following the results of Tables 9 and 10, we can find that the values are greater than 0.05 generally. In the following, we will explain these results in Tables 9 and 10, respectively.

Table 7: Comparisons among the classification accuracies with using SVM.

ID	RAW	APP	MGAPP	RAW	ENT	MGENT	RAW	DIS	MGDIS	RAW	ERR	MGERR
1	0.9471	*0.9422*	**0.9543**	0.9471	*0.9365*	**0.9507**	**0.9471**	0.9471	*0.9436*	**0.9471**	0.9468	*0.9365*
2	**0.4909**	0.4413	*0.4364*	*0.4909*	**0.5527**	0.5345	0.4909	*0.3284*	**0.5873**	0.4909	**0.5020**	*0.4709*
3	**0.9407**	0.9152	*0.9074*	0.9407	*0.9296*	**0.9444**	**0.9407**	0.9407	*0.9222*	**0.9407**	0.9126	*0.9111*
4	**0.9671**	*0.9343*	**0.9671**	**0.9671**	*0.9380*	0.9398	**0.9671**	0.9432	*0.9272*	**0.9671**	*0.9518*	**0.9671**
5	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600
6	*0.8282*	**0.8354**	0.8282	0.8282	*0.8166*	**0.8549**	0.8282	*0.6704*	**0.8282**	0.8282	0.8437	**0.8472**
7	**0.5465**	*0.4084*	0.5446	**0.5465**	*0.4912*	0.5267	**0.5465**	0.5465	*0.5438*	**0.5465**	*0.4967*	0.5446
8	**0.9314**	0.8886	0.9143	**0.9314**	*0.8611*	0.8857	**0.9314**	0.9274	*0.8571*	**0.9314**	*0.8669*	0.8914
9	**0.6909**	0.6204	*0.5709*	**0.6909**	0.6862	*0.6709*	**0.6909**	0.6909	0.6909	0.6909	*0.6342*	**0.7273**
10	**0.8315**	*0.7873*	0.8201	**0.8315**	0.8226	0.8296	**0.8315**	*0.7605*	0.8201	**0.8315**	0.8248	*0.8125*
11	**0.9048**	*0.8914*	0.8952	**0.9048**	*0.9000*	0.9048	**0.9048**	0.9048	0.9048	*0.9048*	0.9048	**0.9143**
12	**0.7480**	*0.7382*	**0.7480**	0.7480	*0.7400*	**0.7500**	0.7480	0.7480	*0.7300*	0.7480	0.7124	*0.7100*
13	**0.8323**	*0.7748*	0.8000	**0.8323**	*0.7600*	0.8194	**0.8323**	*0.8142*	**0.8323**	0.8323	*0.7832*	**0.8323**
14	**0.9889**	0.9800	*0.9778*	**0.9889**	0.9822	*0.9667*	**0.9889**	0.9889	0.9889	**0.9889**	*0.9767*	0.9778
15	**0.5402**	*0.5001*	0.5349	**0.5402**	0.5259	0.5297	**0.5402**	*0.4797*	0.5383	**0.5402**	*0.4444*	*0.4444*
Avg	**0.8032**	*0.7678*	0.7839	**0.8032**	*0.7868*	0.7979	**0.8032**	*0.7700*	0.8013	**0.8032**	*0.7774*	0.7898

The greatest values are in bold, and the smallest values are italics.

Table 8: Comparison among the classification accuracies with using CRAT.

ID	RAW	APP	MGAPP	RAW	ENT	MGENT	RAW	DIS	MGDIS	RAW	ERR	MGERR
1	*0.9122*	0.9154	**0.9365**	*0.9122*	0.9148	**0.9259**	*0.9122*	0.9122	**0.9297**	*0.9122*	**0.9206**	**0.9206**
2	**0.6273**	0.5253	*0.5236*	**0.6273**	*0.5389*	0.5473	**0.6273**	*0.3375*	0.5673	0.6273	*0.5225*	**0.5891**
3	**0.9111**	0.9048	*0.9000*	**0.9111**	*0.9041*	0.9000	**0.9111**	0.9111	*0.9037*	**0.9111**	0.9085	*0.9111*
4	*0.9236*	0.9271	**0.9342**	*0.9236*	0.9386	**0.9450**	0.9236	*0.9056*	**0.9344**	*0.9236*	0.9250	**0.9290**
5	0.8200	**0.8300**	*0.8000*	0.8200	**0.8320**	*0.8200*	0.8200	0.8200	0.8200	*0.8200*	**0.8600**	**0.8600**
6	**0.8396**	*0.8270*	**0.8396**	0.8396	*0.7979*	**0.8509**	**0.8396**	*0.6575*	0.8356	**0.8396**	*0.8377*	0.8280
7	**0.8754**	*0.5849*	0.8404	**0.8754**	0.8223	0.8228	**0.8754**	0.8754	*0.8689*	**0.8754**	*0.8028*	0.8327
8	*0.8629*	**0.8991**	0.8800	0.8629	*0.8297*	**0.8914**	0.8629	0.8651	**0.8800**	0.8629	*0.8411*	**0.8686**
9	**0.7709**	*0.7200*	0.7509	*0.7709*	**0.7989**	0.7909	**0.7709**	0.7709	0.7709	**0.7709**	*0.7644*	0.7673
10	**0.8608**	*0.7872*	0.8239	**0.8608**	*0.8112*	0.8372	**0.8608**	*0.7523*	0.8011	**0.8608**	0.8130	*0.8125*
11	0.9048	*0.8657*	**0.9238**	0.9048	*0.8924*	**0.9143**	0.9048	0.9048	0.9048	*0.9048*	0.9048	**0.9238**
12	**0.6960**	*0.6900*	0.6940	0.6960	*0.6908*	**0.6980**	0.6960	0.6960	*0.6880*	0.6960	0.7094	**0.7240**
13	**0.8129**	0.7329	*0.7161*	**0.8129**	*0.7626*	0.7613	**0.8129**	0.8019	0.8129	**0.8129**	*0.7535*	0.7677
14	*0.9333*	0.9420	**0.9556**	0.9333	**0.9578**	*0.9222*	*0.9333*	0.9333	**0.9444**	*0.9333*	0.9322	0.9333
15	0.5789	*0.5263*	**0.5820**	0.5789	*0.5515*	0.5771	**0.5789**	*0.4739*	0.5746	**0.5789**	*0.4444*	*0.4444*
Avg	**0.8219**	*0.7786*	0.8067	**0.8219**	*0.8029*	0.8136	**0.8219**	*0.7745*	0.8158	**0.8219**	*0.7960*	0.8075

The greatest values are in bold, and the smallest values are italics.

Table 9: Comparisons of the distributions with using SVM.

ID	APP & MGAPP	ENT & MGENT	DIS & MGDIS	ERR & MGERR
1	0.1029	*0.0088*	1.0000	0.9999
2	0.9989	0.9997	*2.9621E − 06*	0.7499
3	1.0000	0.9739	0.9852	1.0000
4	0.8988	0.5226	0.0584	0.9697
5	1.0000	1.0000	1.0000	1.0000
6	0.9967	0.0783	*1.9014E − 07*	0.8709
7	*2.8827E − 36*	*2.0328E − 08*	1.0000	*1.1929E − 09*
8	0.1236	0.9765	0.3849	*0.0043*
9	0.2643	0.9999	1.0000	0.9949
10	0.1887	0.1506	*3.2225E − 13*	0.5573
11	0.9561	0.8719	1.0000	1.0000
12	0.9997	0.9996	0.9999	0.9852
13	0.2446	*0.0261*	0.2749	*0.0392*
14	0.9918	0.8474	1.0000	0.7630
15	*7.8457E − 20*	0.2767	*0.0059*	1.0000
Avg	0.5844	0.5821	0.5806	0.7290

The values less than 0.05 are in italics.

TABLE 10: Comparisons of the distributions with using CART (the values less than 0.05 are italics).

ID	APP & MGAPP	ENT & MGENT	DIS & MGDIS	ERR & MGERR
1	0.0532	$1.9620E-04$	0.9997	0.9912
2	0.1040	1.0000	0.0046	0.9261
3	0.9988	0.9101	1.0000	1.0000
4	0.9933	0.5804	0.1851	0.9985
5	0.9988	0.9997	1.0000	1.0000
6	0.2141	0.0044	$1.6372E-06$	0.9973
7	$2.2575E-52$	$4.1119E-27$	0.9415	$1.8770E-10$
8	0.0253	0.3208	0.9741	0.5315
9	0.9083	1.0000	1.0000	0.9963
10	0.1223	$2.0327E-04$	$6.6803E-06$	0.6861
11	0.4844	0.1356	1.0000	0.9999
12	0.5605	0.9801	1.0000	0.9299
13	0.6600	0.5010	0.2942	0.0936
14	0.9219	0.9374	1.0000	0.8304
15	$8.8496E-26$	$2.6348E-07$	$1.6897E-05$	1.0000
Avg	0.4697	0.4913	0.6266	0.7987

(1) As for those values which are equal toor less than 0.05 in Table 9, it can be observed from Table 7 that the classification accuracies with respect to the multigranularity reduct are greater than the ones in terms of the corresponding single-granularity reducts. Take the data set of "Vertebral Column" (ID: 13) as an example, there are two values no greater than 0.05: in the comparisons of "ENT & MGENT" and "ERR & MGERR." However, the classification accuracies with respect to these reducts (ENT, MGENT, ERR, and MGERR) are 0.7600, 0.8194, 0.7832, and 0.8323. We can find that the values in terms of "MGENT" ("MGERR") are greater than "ENT" ("ERR").

(2) The similar discovery can be observed from Table 10. As for the values which are the same to or less than 0.05, the multigranularity reducts provide greater values of classification accuracies. However, the only exception is the data set "Ionosphere" (ID: 8); in the comparison of "APP & MGAPP," the value of the result is less than 0.05, and the mean values in term of "APP" and "MGAPP" is 0.8991 and 0.8800, respectively.

To sum up, we can conclude that there is not much significant difference between the distributions among the classification accuracies with respect to the single-granularity reducts and the multigranularity reducts. Furthermore, not only the multigranularity reducts can provide better classification accuracies than the single-granularity reducts but also the multigranularity algorithm can improve the time efficiency.

6. Conclusions and Future Perspectives

In this paper, a framework of multigranularity view for computing reduct has been proposed. Different from the traditional algorithm for reduct which uses only one fixed granularity, our algorithm is executed from the multi-granularity view. In the experiment, the finest and the coarsest granularities are employed to realize the multi-granularity framework. Furthermore, to select attributes from the viewpoint of different decision classes, the ensemble strategy is introduced into the traditional fitness function based searching strategy, and the majority mechanism is employed to choose the attribute with the highest frequency with respect to different decision classes. Compared with the single-granularity view for computing reduct, the proposed algorithm can not only reduce the elapsed time but also the reducts derived from the new algorithm can improve the classification accuracies.

The following topics deserve our further investigations:

(1) Both the finest granularity and the coarsest granularity each take up the weight of 1/2 in this paper, and different weights with respect to granularities may be conducted with different requirements in the future investigations.

(2) The weighted average values related to the finest granularity and the coarsest granularity are applied in the multigranularity algorithm, and other more granularities or other more fusion strategies can be considered as the multigranularity view to realize the attribute reduction.

(3) To consider the different requirements of different decision classes, the ensemble selector with the majority voting mechanism is employed to select the suitable attribute in our experiment, and more types of strategies may be investigated among decision classes in the future.

Data Availability

The (UCI) data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. M. Chen, T. R. Li, C. Luo, S.-J. Horng, and G. Wang, "A decision-theoretic rough set approach for dynamic data mining," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 1958–1970, 2015.

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

[3] Z. Pawlak and A. Skowron, "Rough sets: some extensions," *Information Sciences*, vol. 177, no. 1, pp. 28–40, 2007.

[4] X. Zhang, C. Mei, D. Chen, and J. Li, "Multi-confidence rule acquisition and confidence-preserved attribute reduction in interval-valued decision systems," *International Journal of Approximate Reasoning*, vol. 55, no. 8, pp. 1787–1804, 2014.

[5] Q. H. Hu, W. Pedrycz, D. Yu, and J. Lang, "Selecting discrete and continuous features based on neighborhood decision error minimization," *IEEE Transactions on Cybernetics*, vol. 40, no. 1, pp. 137–150, 2010.

[6] J. Dai, W. Wang, H. Tian, and L. Liu, "Attribute selection based on a new conditional entropy for incomplete decision systems," *Knowledge-Based Systems*, vol. 39, pp. 207–213, 2013.

[7] J. Dai, W. Wang, Q. Xu, and H. Tian, "Uncertainty measurement for interval-valued decision systems based on extended conditional entropy," *Knowledge-Based Systems*, vol. 27, pp. 443–450, 2012.

[8] J. Dai, Q. Xu, W. Wang, and H. Tian, "Conditional entropy for incomplete decision systems and its application in data mining," *International Journal of General Systems*, vol. 41, no. 7, pp. 713–728, 2012.

[9] Y. Chen, Y. Xue, Y. Ma, and F. Xu, "Measures of uncertainty for neighborhood rough sets," *Knowledge-Based Systems*, vol. 120, pp. 226–235, 2017.

[10] T. Feng, S.-P. Zhang, and J.-S. Mi, "The reduction and fusion of fuzzy covering systems based on the evidence theory," *International Journal of Approximate Reasoning*, vol. 53, no. 1, pp. 87–103, 2012.

[11] Y.-Q. Yao, J.-S. Mi, and Z.-J. Li, "Attribute reduction based on generalized fuzzy evidence theory in fuzzy decision systems," *Fuzzy Sets and Systems*, vol. 170, no. 1, pp. 64–75, 2011.

[12] J. Li, C. Mei, W. Xu, and Y. Qian, "Concept learning via granular computing: a cognitive viewpoint," *Information Sciences*, vol. 298, pp. 447–467, 2015.

[13] L. A. Zadeh, "Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems," *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, vol. 2, no. 1, pp. 23–25, 1998.

[14] L. A. Zadeh, "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 111–127, 1997.

[15] P. Zhu and Q. Hu, "Adaptive neighborhood granularity selection and combination based on margin distribution optimization," *Information Sciences*, vol. 249, pp. 1–12, 2013.

[16] P. Zhu, Q. Hu, W. Zuo, and M. Yang, "Multi-granularity distance metric learning via neighborhood granule margin maximization," *Information Sciences*, vol. 282, pp. 321–331, 2014.

[17] Y. Kudo, T. Murai, and S. Akama, "A granularity-based framework of deduction, induction, and abduction," *International Journal of Approximate Reasoning*, vol. 50, no. 8, pp. 1215–1226, 2009.

[18] S. U. Kumar and H. H. Inbarani, "A novel neighborhood rough set based classification approach for medical diagnosis," *Procedia Computer Science*, vol. 47, pp. 351–359, 2015.

[19] M.-W. Shao, Y. Leung, X.-Z. Wang, and W.-Z. Wu, "Granular reducts of formal fuzzy contexts," *Knowledge-Based Systems*, vol. 114, no. 15, pp. 156–166, 2016.

[20] W.-Z. Wu and Y. Leung, "Optimal scale selection for multi-scale decision tables," *International Journal of Approximate Reasoning*, vol. 54, no. 8, pp. 1107–1129, 2013.

[21] W.-Z. Wu, Y. Qian, T.-J. Li, and S.-M. Gu, "On rule acquisition in incomplete multi-scale decision tables," *Information Sciences*, vol. 378, pp. 282–302, 2017.

[22] H. Zhao, P. Wang, and Q. Hu, "Cost-sensitive feature selection based on adaptive neighborhood granularity with multi-level confidence," *Information Sciences*, vol. 366, pp. 134–149, 2016.

[23] S.-M. Gu and W.-Z. Wu, "On knowledge acquisition in multi-scale decision systems," *International Journal of Machine Learning and Cybernetics*, vol. 4, no. 5, pp. 477–486, 2013.

[24] Y. Yao, "A triarchic theory of granular computing," *Granular Computing*, vol. 1, no. 2, pp. 145–157, 2016.

[25] Y. Lin, J. Li, P. Lin, G. Lin, and J. Chen, "Feature selection via neighborhood multi-granulation fusion," *Knowledge-Based Systems*, vol. 67, pp. 162–168, 2014.

[26] Y. Qian, J. Liang, Y. Yao, and C. Dang, "MGRS: a multi-granulation rough set," *Information Sciences*, vol. 180, no. 6, pp. 949–970, 2010.

[27] Y. Y. Yao, "Three-way decision and granular computing," *International Journal of Approximate Reasoning*, vol. 103, pp. 107–123, 2018.

[28] K. Liu, X. Yang, H. Fujita, D. Liu, X. Yang, and Y. Qian, "An efficient selector for multi-granularity attribute reduction," *Information Sciences*, vol. 505, pp. 457–472, 2019.

[29] D. G. Chen and S. Y. Zhao, "Local reduction of decision system with fuzzy rough set," *Fuzzy Sets and Systems*, vol. 161, no. 13, pp. 1871–1883, 2010.

[30] X. Yang and Y. Yao, "Ensemble selector for attribute reduction," *Applied Soft Computing*, vol. 70, pp. 1–11, 2018.

[31] J. Song, E. C. C. Tsang, D. Chen, and X. Yang, "Minimal decision cost reduct in fuzzy decision-theoretic rough set model," *Knowledge-Based Systems*, vol. 126, pp. 104–112, 2017.

[32] H. Chen, T. Li, X. Fan, and C. Luo, "Feature selection for imbalanced data based on neighborhood rough sets," *Information Sciences*, vol. 483, pp. 1–20, 2019.

[33] J. Bai, K. Xia, Y. Lin, and P. Wu, "Attribute reduction based on consistent covering rough set and its application," *Complexity*, vol. 2017, Article ID 8986917, 9 pages, 2017.

[34] B. Sun, W. Ma, and Y. Qian, "Multigranulation fuzzy rough set over two universes and its application to decision making," *Knowledge-Based Systems*, vol. 123, pp. 61–74, 2017.

[35] X. Qi, J. Zhang, and C. Liang, "Multiple attributes group decision-making approaches based on interval-valued dual hesitant fuzzy unbalanced linguistic set and their applications," *Complexity*, vol. 2018, Article ID 3172716, 22 pages, 2018.

[36] Y. Qian, X. Liang, Q. Wang et al., "Local rough set: a solution to rough data analysis in big data," *International Journal of Approximate Reasoning*, vol. 97, pp. 38–63, 2018.

[37] Z. Jiang, X. Yang, H. Yu, D. Liu, P. Wang, and Y. Qian, "Accelerator for multi-granularity attribute reduction," *Knowledge-Based Systems*, vol. 177, pp. 145–158, 2019.

[38] J. Liang and Z. Shi, "The information entropy, rough entropy and knowledge granulation in rough set theory," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 12, no. 1, pp. 37–46, 2004.

[39] J. Liang, J. Wang, and Y. Qian, "A new measure of uncertainty based on knowledge granulation for rough sets," *Information Sciences*, vol. 179, no. 4, pp. 458–470, 2009.

[40] W. Wei and J. Liang, "Information fusion in rough set theory: an overview," *Information Fusion*, vol. 48, pp. 107–118, 2019.

[41] S. Xu, X. Yang, H. Yu, D.-J. Yu, J. Yang, and E. C. C. Tsang, "Multi-label learning with label-specific feature reduction," *Knowledge-Based Systems*, vol. 104, pp. 52–61, 2016.

[42] Y. Gao, X. Chen, X. Yang, and P. Wang, "Neighborhood attribute reduction: a multicriterion strategy based on sample selection," *Information*, vol. 9, no. 11, p. 282, 2018.

[43] Q. Hu, X. Che, L. Zhang, D. Zhang, M. Guo, and D. Yu, "Rank entropy-based decision trees for monotonic classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2052–2064, 2012.

[44] X. Zhang, C. Mei, D. Chen, and J. Li, "Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy," *Pattern Recognition*, vol. 56, no. 1, pp. 1–15, 2016.

[45] C. Z. Wang, Q. H. Hu, X. Z. 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, 2018.

[46] Q. Hu, D. Yu, and Z. Xie, "Neighborhood classifiers," *Expert Systems with Applications*, vol. 34, no. 2, pp. 866–876, 2008.

[47] X. Yang, S. Liang, H. Yu, S. Gao, and Y. Qian, "Pseudo-label neighborhood rough set: measures and attribute reductions," *International Journal of Approximate Reasoning*, vol. 105, pp. 112–129, 2019.

[48] X. Y. Jia, Y. Rao, L. Shang, and T. Li, "Similarity-based attribute reduction in rough set theory: a clustering perspective," *International Journal of Machine Learning and Cybernetics*, vol. 1, pp. 1–14, 2019.

[49] Y. Liu, W. L. Huang, Y. L. Jiang, and Z. Y. Zeng, "Quick attribute reduct algorithm for neighborhood rough set model," *Information Sciences*, vol. 271, no. 1, pp. 65–81, 2014.

[50] 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.

[51] X. Yang, Y. Qi, X. Song, and J. Yang, "Test cost sensitive multigranulation rough set: model and minimal cost selection," *Information Sciences*, vol. 250, pp. 184–199, 2013.

[52] X. Yang, M. Zhang, H. Dou, and J. Yang, "Neighborhood systems-based rough sets in incomplete information system," *Knowledge-Based Systems*, vol. 24, no. 6, pp. 858–867, 2011.

[53] Y. Yao and X. Zhang, "Class-specific attribute reducts in rough set theory," *Information Sciences*, vol. 418-419, pp. 601–618, 2017.

[54] X. Jia, L. Shang, B. Zhou, and Y. Yao, "Generalized attribute reduct in rough set theory," *Knowledge-Based Systems*, vol. 91, pp. 204–218, 2016.

[55] Y. Yao, Y. Zhao, and J. Wang, "On reduct construction algorithms," in *Lecture Notes in Computer Science*, vol. 1, pp. 100–117, Springer, Berlin, Germany, 2008.

[56] K. Liu, X. Yang, H. Yu, J. Mi, P. Wang, and X. Chen, "Rough set based semi-supervised feature selection via ensemble selector," *Knowledge-Based Systems*, vol. 165, pp. 282–296, 2019.

[57] J. Li, C. Huang, J. Qi, Y. Qian, and W. Liu, "Three-way cognitive concept learning via multi-granularity," *Information Sciences*, vol. 378, pp. 244–263, 2017.

[58] 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.

[59] K. Y. Liu, J. J. Song, W. D. Zhang, and X. B. Yang, "Alleviating over-fitting in attribute reduction: an early stopping strategy," in *Proceedings of the 2018 International Conference on Wavelet Analysis and Pattern Recognition*, pp. 190–195, Chengdu, China, July 2018.

[60] R. Wang, W. Li, R. Li, and L. Zhang, "Automatic blur type classification via ensemble SVM," *Signal Processing: Image Communication*, vol. 71, pp. 24–35, 2019.

[61] H. Yu, C. Mu, C. Sun, W. Yang, X. Yang, and X. Zuo, "Support vector machine-based optimized decision threshold adjustment strategy for classifying imbalanced data," *Knowledge-Based Systems*, vol. 76, pp. 67–78, 2015.

[62] H. Yu, S. Changyin, X. Yang, S. Zheng, and H. Zou, "Fuzzy support vector machine with relative density information for classifying imbalanced data," *IEEE Transactions on Fuzzy Systems*, vol. 99, pp. 1–15, 2019.

[63] I. Carmela, M. Ariab, D. A. Antonio, and R. Siciliano, "Informative trees by visual pruning," *Expert Systems with Applications*, vol. 127, pp. 228–240, 2019.