Fuzzy Systems and Data Mining VI
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA200703

Neighborhood Based Multi-Granularity Attribute Reduction: An Acceleration Approach

Jingjing SONG [a,b,1], Huili DOU [a], Xiansheng RAO [a], Xiaojing LUO [a] and Xuan YAN [c]
[a] *School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China*
[b] *Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou, 363000, China*
[c] *School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China*

Abstract. As a feature selection technique in rough set theory, attribute reduction has been extensively explored from various viewpoints especially the aspect of granularity, and multi-granularity attribute reduction has attracted much attention. Nevertheless, it should be pointed out that multiple granularities require to be considered simultaneously to evaluate the significance of candidate attribute in the corresponding process of computing reduct, which may result in high elapsed time of searching reduct. To alleviate such a problem, an acceleration strategy for neighborhood based multi-granularity attribute reduction is proposed in this paper, which aims to improve the computational efficiency of searching reduct. Our proposed approach is actually realized through the positive approximation mechanism, and the processes of searching qualified attributes are executed through evaluating candidate attributes over the gradually reduced sample space rather than all samples. The experimental results over 12 UCI data sets demonstrate that the acceleration strategy can provide superior performance to the naive approach of deriving multi-granularity reduct in the elapsed time of computing reduct without generating different reducts.

Keywords. Acceleration approach, attribute reduction, granular conputing, multi-granularity, neighborhood rough set

1. Introduction

Attribute reduction [1,2,3,4,5], as a rough set based feature selection technology, has been widely investigated from various perspectives, and it has also been applied to many fields such as pattern recognition [6], decision analysis[7,8,9], data mining [10] and machine learning [11,12,13]. This is mainly because the data collected in real-world applications may contain redundant and irrelative attributes, these attributes may deteriorate the performance of learning algorithms [14,15], attribute reduction can effectively

[1]Corresponding Author: Jingjing Song, School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China; E-mail: songjingjing108@163.com

remove these attributes from data through searching a qualified reduct satisfying the intended constraint, and further reduce the dimensionality of data. Moreover, it is worth mentioning that Rough Set Theory (RST) has its inherent superiority compared with other methods. For instance, it is effective in handling uncertainty without the requirement of prior information.

Notably, the intended constraint defined in attribute reduction is generally constructed over the given measure such as approximation quality [16] or condition entropy [17], and the corresponding derived measure value is highly related to the result of information granulation [18,19,20] over the universe from the standpoint of Granular Computing (GrC) [21,22]. For example, if the neighborhood rough set [23] and measure approximation quality are considered, and given a radius, then the obtained neighborhoods of samples can be regarded as the neighborhood information granules, i.e., the result of information over the universe. Therefore, a smaller size of neighborhood, which indicates the corresponding result of information granulation is finer, may generate higher measure value of approximation quality; contrarily, a greater size of neighborhood, which implies the corresponding result of information granulation is coarser, may lead to lower measure value of approximation quality.

Presently, the concept of granularity has been used for characterizing the level of information granulation in GrC. And the corresponding level of granularity can reflect the discrimination ability related to the result of information granulation. With a careful reviewing of previous research, it is easy to reveal the intended constraint is generally constructed based on one and only one fixed granularity in most attribute reduction approaches, which are referred to as single granularity based attribute reduction in the context of this paper. However, as what has been illustrated in References [24,25,26], there are some inherent limitations in single granularity based attribute reduction. For example, single granularity based attribute reduction may fail to select attributes from multi-level or multi-view [24]; single granularity based attribute reduction may not provide the higher adaptability of the derived reduct for the problem of granularity diversity [26]. Therefore, various definitions of multi-granularity attribute reduction have been proposed for different practical requirements. Notably, Liu et al. [26] not only put forward a general definition for multi-granularity attribute reduction, but also developed the corresponding algorithm of searching reduct.

However, it is worth noting that in multi-granularity attribute reduction, multiple different granularities are required to be considered simultaneously to calculate the measure value, which will be further employed for evaluating the significance of candidate attributes. Immediately, qualified attributes are selected and added into the temporary attribute set until the intended constraints are satisfied. Obviously, such process of searching reduct is time-consuming due to the simultaneous consideration of multiple different granularities especially the size of samples is large. In view of this, to reduce the computational time of obtaining reduct, an acceleration approach is proposed for neighborhood based multi-granularity attribute reduction in neighborhood rough set. Our proposed acceleration approach is mainly designed through the positive approximation mechanism [15]. Specifically, to search the qualified candidate attributes, the significance of candidate attribute can be evaluated over the gradually reduced sample space rather than all samples. Accordingly, the decreasing of sample space may be conducive to the less elapsed time of calculating measure value and selecting attributes. Consequently,

the computations of deriving reduct may be saved along with the time consumption of searching reduct.

The remainder of this paper is organized as follows. The basic notions related to neighborhood rough set and multi-granularity attribute reduction will be briefly reviewed in Section 2. The acceleration approach for neighborhood based multi-granularity attribute reduction will be presented in Section 3. Comparative experiments will be conducted to verify the performance of the proposed acceleration strategy in Section 4. Section 5 will conclude this paper and present future perspectives.

2. Preliminary knowledge

Generally speaking, the obtaining of granularity can be realized through various strategies [26], one of which is the parameter based granularity. Therefore, to simplify our discussion, the neighborhood rough set is employed in the following paper. The main reason can be summarized as the following two points: 1) neighborhood rough set is effective in analysing and handling continuous data and even mixed data; 2) neighborhood rough set intuitively determines the multi-granularity structure based on the concept of neighborhood [23] with considering multiple different radii.

2.1. Neighborhood rough set

From the viewpoint of rough set theory [16], a decision system can be described as a pair such that $DS =< U, A \cup \{d\} >$, in which U is the universe, i.e., the nonempty finite set of samples, A is the condition attribute set used for describing the sample, and d is the decision attribute which indicates the true label of sample. Following the label of sample, the equivalence relation for information granulation can be derived, which is denoted as $\mathrm{IND}_d = \{(x,y) \in U \times U : d(x) = d(y)\}$. Based on such equivalence relation, a partition $U/\mathrm{IND}_d = \{X_1, X_2, \ldots, X_q\}$ over the universe can be induced, in which $X_p (X_p \in U/\mathrm{IND}_d)$ is the p-th decision class.

It should be noticed that the classical rough set is only useful in analysing and handling nominal data. However, numerical data and mixed data are ubiquitous in real-world applications. Therefore, various generalizations [23,27,28,29,30] of such model have been proposed. Neighborhood rough set has been favoured and employed by many researchers, because it can analyse and handle continuous data and even mixed data directly.

Definition 1 *[23] Given a decision system DS and the neighborhood radius δ, $\forall B \subseteq A$, the neighborhood of sample x is:*

$$N_B^{\delta}(x) = \{y \in U : \Delta_B(x,y) \leq \delta\}, \tag{1}$$

in which $\Delta_B(x,y)$ is the distance between samples x and y over the condition attribute set B.

From the viewpoint of GrC, the result of information granulation is actually the neighborhoods of all samples over the universe in neighborhood rough set, i.e., all the neighborhood information granules. Therefore, different scales of neighborhood of sam-

ple can generate different results of information granulation, and then different granularities can be obtained. That is, the granularity can be reflected by the derived neighborhoods or the used neighborhood radius. This is mainly because a smaller radius may generate a smaller size of neighborhood which indicates a finer granularity, while a larger radius may determine a greater size of neighborhood which implies a coarser granularity. From this perspective, a set of different radii can determine a set of different granularities.

Following the definition of neighborhood of sample, the neighborhood relation over the universe can be induced, and it can be denoted as $\mathrm{N}_B^\delta = \{(x,y) \in U \times U : \Delta_B(x,y) \leq \delta\}$.

Definition 2 *[23] Given a decision system DS and the neighborhood radius δ, $\forall B \subseteq A$, the neighborhood lower and upper approximations of X_p with respect to B are:*

$$\underline{\mathrm{N}_B^\delta}(X_p) = \{x \in U : N_B^\delta(x) \subseteq X_p\}, \tag{2}$$

$$\overline{\mathrm{N}_B^\delta}(X_p) = \{x \in U : N_B^\delta(x) \cap X_p \neq \emptyset\}. \tag{3}$$

The pair $[\underline{\mathrm{N}_B^\delta}(X_p), \overline{\mathrm{N}_B^\delta}(X_p)]$ is referred to as a neighborhood rough set of X_p with respect to B.

Definition 3 *[23] Given a decision system and the neighborhood radius δ, $\forall B \subseteq A$, the approximation quality of d with respect to B is*

$$\gamma_B^\delta(U,d) = \frac{|\mathrm{POS}_B^\delta(U,d)|}{|U|}, \tag{4}$$

in which $\mathrm{POS}_B^\delta(U,d) = \bigcup_{p=1}^{q} \underline{\mathrm{N}_B^\delta}(X_p)$ is the positive region of d with respect to B, and $|X|$ indicates the cardinality of the set X.

2.2. Multi-granularity attribute reduction

Multi-granularity attribute reduction implies that multiple granularities are generally considered for the constructing of intended constraints. Accordingly, a general form of definition proposed by Liu et al. [26] is shown as following Definition 4.

Definition 4 *[26] Given the decision system DS and a set of granularities $\mathbb{MG} = \{\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_n\}$, assuming that $C_\varphi^{\mathbb{MG}}$ is a multi-granularity constraint with regard to measure φ, then $\forall B \subseteq A$, B is a multi-granularity reduct if and only if*

1. *B satisfies $C_\varphi^{\mathbb{MG}}$;*
2. *$\forall C \subset B$, C does not satisfy $C_\varphi^{\mathbb{MG}}$.*

Following Definition 4, it can be observed that different from the constraint in single granularity based attribute reduction, $C_\varphi^{\mathbb{MG}}$ is a multi-granularity constraint constructed over a set of multiple different granularities with respect to the considered measure φ.

Actually, the multi-granularity constraint $C_\varphi^{\mathbb{MG}}$ shown in Definition 4 is realized based on the derived multi-granularity measure value with respect to the considered

measure φ. For example, higher measure value is expected based on the considered measure φ, such as approximation quality, if the corresponding measure value with regard to B and granularity $\mathbb{G}_m$ is denoted as $\varphi^{\mathbb{G}_m}(B)$, then such multi-granularity measure value $\varphi^{\mathbb{MG}}(B)$ in constraints can be induced through fusing the measure values in $\{\varphi^{\mathbb{G}_1}(B), \varphi^{\mathbb{G}_2}(B), \ldots, \varphi^{\mathbb{G}_n}(B)\}$. Immediately, the corresponding constraint can be expressed as "$\varphi^{\mathbb{MG}}(B) \geq \varphi^{\mathbb{MG}}(A)$".

Note that a set of multiple granularities $\mathbb{MG} = \{\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_n\}$ shown in Definition 4 can be intuitively determined by a set of radii $\mathbb{R} = \{\delta_1, \delta_2, \ldots, \delta_n\}$, because the granularity is highly correlated with the considered parameter (the neighborhood radius) in neighborhood rough set. Therefore, in neighborhood rough set, if a set of different radii $\mathbb{R}$ and approximation quality are considered, then the multi-granularity constraint $C_\varphi^{\mathbb{MG}}$ shown in Definition 4 can be denoted as "$\gamma_B^{\mathbb{R}}(U,d) \geq \gamma_A^{\mathbb{R}}(U,d)$", where $\gamma_B^{\mathbb{R}}(U,d)$ is a multi-granularity measure value of approximation quality derived by fusing the corresponding measure values related to multiple radii.

Following Definition 4, how to search the corresponding reduct should be immediately addressed. Note that the heuristic searching algorithm has been widely used for the deriving of reduct, and the significance of candidate attribute is generally regarded as the heuristic information in such process. Based on the neighborhood rough set and measure of approximation quality, Algorithm 1 shows the detailed process of computing multi-granularity reduct.

Algorithm 1: The process of computing multi-granularity reduct.

Inputs: Decision system DS, a set of radii $\mathbb{R}$ with ascending order.
Outputs: One multi-granularity reduct B.
1. $\forall \delta_m \in \mathbb{R}$, calculate the measure value $\gamma_A^{\delta_m}(U,d)$;//$1 \leq m \leq n$;
2. Compute the fused measure value $\gamma_A^{\mathbb{R}}(U,d)$ in terms of the obtained measure values in Step 1;
3. $B \leftarrow \emptyset$;
4. Repeat
1) $\forall a \in A - B$, calculate the measure value $\phi(B \cup \{a\}) = \gamma_{B\cup\{a\}}^{\delta_1}(U,d) + \gamma_{B\cup\{a\}}^{\delta_n}(U,d)$;
2) Select a suitable b such that $b = \arg\max\{\phi(B \cup \{a\}) : \forall a \in A - B\}$;
3) $B \leftarrow B \cup \{b\}$;
4) Calculate the measure value $\gamma_B^{\mathbb{R}}(U,d)$;
Until $\gamma_B^{\mathbb{R}}(U,d) \geq \gamma_A^{\mathbb{R}}(U,d)$
5. Return A.

In Algorithm 1, the following two points should be concerned.

1. The fused measure value $\gamma_A^{\mathbb{R}}(U,d)$ is actually obtained by fusing the multiple measure values with regard to a set of radii, and such fusion operation can be realized through various approaches. For instance, if the arithmetic mean is considered to conduct such fusion, then the fused measure value can be denoted as $\gamma_A^{\mathbb{R}}(U,d) = \frac{1}{n}\Sigma_{m=1}^{n}\gamma_A^{\delta_m}(U,d)$; if the harmonic mean is considered to perform such fusion, then the fused measure value can be formulated as $\gamma_A^{\mathbb{R}}(U,d) = (\frac{1}{n}\Sigma_{m=1}^{n}\frac{1}{\gamma_A^{\delta_m}(U,d)})^{-1}$.

2. The selecting of qualified attributes are executed based on the evaluation of attribute significance over the minimal radius and maximal radius simultaneously, and such two radii can determine the finest and coarsest granularities, respectively.

Following Algorithm 1, two representative radii, i.e., the minimal and maximal radii, should be concerned simultaneously in evaluating candidate attributes. Moreover, the deriving of the measure value in the intended constraint requires to fuse measure values of approximation quality over multiple different radii. It follows that the computational cost of obtaining multi-granularity reduct may be expensive if the volume of data is great.

3. Acceleration approach for neighborhood based multi-granularity attribute reduction

From the above discussions, it is trivial to observe that multiple granularities require to be considered in multi-granularity attribute reduction simultaneously, and then the computing of reduct may be a time-consuming process. In view of this, an acceleration approach will be proposed in the following, which aims to improve the computational efficiency of deriving the neighborhood based multi-granularity reduct.

Such acceleration strategy is mainly realized based on the mechanism of positive approximation [15]. Specifically, with the growing of number of used attributes in neighborhood rough set, smaller size of neighborhood will be induced, which may result in greater size of positive region, and this is determined by the monotonicity of neighborhood rough set. Apparently, the size of previously derived positive region is smaller than that of currently obtained positive region, and the former will be contained in the latter in the iterations of searching qualified attributes. Consequently, the samples in previously obtained positive region will not be participated in next iteration for evaluating candidate attributes. That is, the candidate attributes can be evaluated over the reduced sample space rather than all samples. It follows that the computation of deriving reduct may be decreased along with the elapsed time. Algorithm 2 shows the detailed process of quick computing multi-granularity reduct.

Algorithm 2: The process of quick computing multi-granularity reduct.

Inputs: Decision system DS, a set of radii $\mathbb{R}$ with ascending order.
Outputs: One multi-granularity reduct B.
1. $\forall \delta_m \in \mathbb{R}$, calculate the measure value $\gamma_A^{\delta_m}(U,d)$; //$1 \leq m \leq n$
2. Compute the fused measure value $\gamma_A^{\mathbb{R}}(U,d)$ in terms of the obtained measure values in Step 1;
3. $B \leftarrow \emptyset$
4. Repeat
1) Compute the positive regions $\mathrm{POS}_B^{\delta_1}(U,d)$ and $\mathrm{POS}_B^{\delta_n}(U,d)$ with respect to δ_1 and δ_n, respectively;//$\mathrm{POS}_{\emptyset}^{\delta_m}(U,d) = \emptyset$
2) $U_1 \leftarrow U - \mathrm{POS}_B^{\delta_1}(U,d)$, $U_n \leftarrow U - \mathrm{POS}_B^{\delta_n}(U,d)$;
3) $\forall a \in A - B$, calculate the measure value $\phi(B \cup \{a\}) = \gamma_{B\cup\{a\}}^{\delta_1}(U_1,d) + \gamma_{B\cup\{a\}}^{\delta_n}(U_n,d)$;

4) Select a suitable b such that $b = \arg\max\{\phi(B \cup \{a\}) : \forall a \in A - B\}$;
5) $B \leftarrow B \cup \{b\}$;
6) Calculate the measure value $\gamma_B^{\mathbb{R}}(U,d)$
Until $\gamma_B^{\mathbb{R}}(U,d) \geq \gamma_A^{\mathbb{R}}(U,d)$
5. Return A.

In Algorithm 2, it is worth noting that to search the qualified attributes, the selecting of candidate attributes will be executed over the minimal and maximal radii simultaneously. Therefore, the positive regions with respect to such two radii require to be calculated beforehand, respectively, and then the candidate attributes can be evaluated over the reduced sample space.

Moreover, based on the processes of Algorithms 1 and 2, it is easy to know that the sample space of Algorithm 2 is gradually reduced in the iteration of searching qualified attributes, while the sample space of Algorithm 1 is the whole universe in the iteration of searching qualified attributes. Accordingly, compared with the naive process of computing multi-granularity reduct, i.e., Algorithm 1, the redundant computations of deriving multi-granularity reduct may be reduced through the acceleration strategy, i.e., Algorithm 2. Therefore, the time consumption of deriving corresponding reduct may be also decreased.

4. Experiments

4.1. Data sets

To validate the effectiveness of the proposed acceleration approach, 12 public UCI data sets have been used for conducting experiments, the detailed description of data sets is shown in Table 1. All data sets have been normalized by column before conducting experiments.

Table 1. Data sets used in experiments

ID	Data sets	Samples	Attributes	Decision classes
1	Amphetamines	1885	12	7
2	Connectionist Bench	990	13	11
3	Forest Type Mapping	523	27	4
4	Libras Movement	360	90	15
5	Lymphography	98	18	3
6	Optical Recognition of Handwritten Digits	5620	64	10
7	Page-blocks	5473	10	5
8	Statlog (German Credit)	1000	24	2
9	Statlog (Heart)	270	13	2
10	Urban Land Cover	675	147	9
11	Waveform	5000	40	3
12	Wine Quality	4898	11	7

4.2. Experimental setup

In the following experiments, a set of ascending ordered radii $\mathbb{R}$ will be appointed. Note that the same set of radii may be not suitable for all data sets. Therefore, for different data sets, 10 different neighborhood radii has been selected, and the employed set of radii is the same to what has been used in Reference [31], and the details of determining the suitable radii can be observed in the corresponding illustration of References [31]. The detailed values of radii used in our experiments are shown in Table 2.

Table 2. Detailed values of radii for data sets

ID	10 used radii
1	0.021, 0.042, 0.063, 0.084, 0.105, 0.126, 0.147, 0.168, 0.199, 0.210
2	0.021, 0.042, 0.063, 0.084, 0.105, 0.126, 0.147, 0.168, 0.199, 0.210
3	0.025, 0.050, 0.075, 0.100, 0.125, 0.150, 0.175, 0.200, 0.225, 0.250
4	0.054, 0.108, 0.162, 0.216, 0.270, 0.324, 0.378, 0.432, 0.486, 0.540
5	0.037, 0.074, 0.111, 0.148, 0.185, 0.222, 0.259, 0.296, 0.333, 0.370
6	0.047, 0.094, 0.141, 0.188, 0.235, 0.282, 0.329, 0.376, 0.423, 0.470
7	0.014, 0.028, 0.042, 0.056, 0.070, 0.084, 0.098, 0.112, 0.126, 0.140
8	0.041, 0.082, 0.123, 0.164, 0.205, 0.246, 0.287, 0.328, 0.369, 0.410
9	0.045, 0.090, 0.135, 0.180, 0.225, 0.270, 0.315, 0.360, 0.405, 0.450
10	0.035, 0.070, 0.105, 0.140, 0.175, 0.210, 0.245, 0.280, 0.315, 0.350
11	0.047, 0.094, 0.141, 0.188, 0.235, 0.282, 0.329, 0.376, 0.423, 0.470
12	0.011, 0.022, 0.033, 0.044, 0.055, 0.066, 0.077, 0.088, 0.099, 0.110

To test the performance of the derived reduct, 5-fold cross-validation will be used. Note that the comparative experimental results are the average values of 5-fold cross-validation output results in our experiments. Moreover, Euclidean distance has been used in the computing of the distance between arbitrary two samples, and the operator arithmetic mean will be used for realizing the fusion operation in Algorithms 1 and 2. To avoid the constraints in Algorithms 1 and 2 are so strict that redundant attributes may not be removed, the tolerant threshold ε has been considered to realize the approximated reduct [32]. Specifically, both the constraints in Algorithms 1 and 2 can be reset to be "$\gamma_A^{\mathbb{R}}(U,d) - \gamma_B^{\mathbb{R}}(U,d) \leq \varepsilon \cdot \gamma_A^{\mathbb{R}}(U,d)$". Two different values of ε will be used in the following experiments, and they are 0.05 and 0.10, respectively.

4.3. Comparisons among the time consumption of computing reduct

The computational time for obtaining reduct among Algorithms 1 and 2 will be mainly compared in this experiment. The detailed elapsed time of obtaining multi-granularity reduct is shown in Table 3, where the lower values are highlighted in underline.

Table 3. Time consumption (seconds) for computing multi-granularity reduct

	$\varepsilon = 0.05$		$\varepsilon = 0.10$	
ID	**Algorithm 1**	**Algorithm 2**	**Algorithm 1**	**Algorithm 2**
1	7.2617	4.8934	6.7875	4.6246
2	1.8828	1.3655	1.8055	1.2769
3	2.1861	1.5156	1.7224	1.3036
4	3.3829	2.4006	2.4204	1.8659
5	0.0663	0.0639	0.0597	0.0490
6	423.9869	301.8563	400.5364	291.9283
7	31.2085	23.7398	29.3262	23.1574
8	4.6161	2.6439	4.0404	2.5684
9	0.1810	0.1526	0.1629	0.1421
10	16.0863	11.9105	13.9133	10.7305
11	159.1719	135.6123	149.6678	129.4193
12	44.1671	29.1053	40.8465	28.7957

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

1. Compared with Algorithm 1 of computing multi-granularity reduct, the proposed acceleration strategy, i.e., Algorithm 2, can offer the superior performance to Algorithm 1 in the computational time of deriving reduct. From this point of view, the proposed acceleration strategy does help to decrease the time consumption of deriving reduct.
2. If the value of used threshold increases, then the time consumption of deriving multi-granularity reduct may be decreased. This is mainly because when the threshold value increases, then the relevant constraint will become looser, and to search the reduct which satisfies such contraint, fewer qualified attributes may be required. Accordingly, the computational time for obtaining reduct may be saved.

4.4. Comparisons among lengths of the obtained reducts

In this experiments, the lengths of derived multi-granularity reducts in terms of Algorithm 1 and Algorithm 2 will be compared. Note that the compared results of lengths are the mean values of 5-fold cross-validation experimental results. The detailed results are shown in Table 4.

Table 4. Lengths of derived multi-granularity reducts

	$\varepsilon = 0.05$		$\varepsilon = 0.10$	
ID	**Algorithm 1**	**Algorithm 2**	**Algorithm 1**	**Algorithm 2**
1	9.0000	9.0000	8.4000	8.4000
2	8.0000	8.0000	7.4000	7.4000
3	16.6000	16.6000	13.2000	13.2000
4	16.2000	16.2000	12.0000	12.0000
5	7.4000	7.4000	6.8000	6.8000
6	12.2000	12.2000	11.6000	11.6000
7	5.0000	5.0000	4.6000	4.6000
8	11.6000	11.6000	10.0000	10.0000
9	10.0000	10.0000	9.0000	9.0000
10	11.6000	11.6000	10.2000	10.2000
11	20.0000	20.0000	18.2000	18.2000
12	9.0000	9.0000	8.4000	8.4000

With a deep investigation of Table 4, it is not difficult to observe the following.

1. Compared with Algorithm 1 of computing multi-granularity reduct, our proposed acceleration approach can generate the same length of reduct. Such observation implies that our proposed acceleration approach may contribute to the deriving of the same reduct as that of Algorithm 1.
2. With the increasing of value of employed threshold, the lengths of the obtained reducts may be reduced. The explanation for such observation result is that when the threshold value increases, the predefined constraint may not be so strict. Accordingly, to derived the qualified multi-granularity reduct, fewer attributes are required. Consequently, the attributes in reduct may be reduced as well as the length of derived reduct. Such reason is consistent with what has been illustrated in previous subsection.

4.5. Comparisons among different reducts in classification performance

The classification performance with respect to the different reducts, which are obtained through Algorithms 1 and 2 will be compared in this experiments. Note that SVM (LIBSVM [33]) classifier is used for estimating the classification performance of reduct over the testing set. Accordingly, the average values of classification accuracies obtained by cross-validation experimental results will be listed, and the detailed results are shown in Table 5.

Table 5. Classification accuracies of multi-granularity reducts

	$\varepsilon = 0.05$		$\varepsilon = 0.10$	
ID	**Algorithm 1**	**Algorithm 2**	**Algorithm 1**	**Algorithm 2**
1	0.5194	0.5194	0.5194	0.5194
2	0.5343	0.5343	0.5131	0.5131
3	0.8606	0.8606	0.8567	0.8567
4	0.4556	0.4556	0.4444	0.4444
5	0.8037	0.8037	0.8037	0.8037
6	0.8909	0.8909	0.8794	0.8794
7	0.9357	0.9357	0.9344	0.9344
8	0.7530	0.7530	0.7540	0.7540
9	0.8333	0.8333	0.8074	0.8074
10	0.7763	0.7763	0.7719	0.7719
11	0.8654	0.8654	0.8584	0.8584
12	0.5343	0.5343	0.5300	0.5300

Based on Table 5, it is not difficult to observe the following.

1. The classification accuracies obtained by using Algorithm 2 are same with those obtained through using Algorithm 1. Such result shows that the proposed acceleration strategy may generate the same reduct with Algorithm 1.
2. The increasing of value of threshold may lead to the decreasing of classification accuracy of obtained reduct over most data sets. This is mainly because with the increasing of value of the considered threshold, the related constraint may be looser, and then the derived reduct may contain fewer attributes. Note that the reduct of fewer attributes generally offers the poor classification performance. It follows that the classification accuracies of reducts may be decreased with the increasing of value of the used threshold.

Based on the results shown in Tables 3-5, we can easily conclude that the acceleration approach can provide superior performance to Algorithm 1 in term of the time efficiency of computing multi-granularity reduct, and the generated reducts are the same with those obtained by using Algorithm 1.

5. Conclusions and future perspectives

Multi-granularity attribute reduction is mainly investigated in this paper, and to search the corresponding reduct, multiple different granularities are generally required to be considered simultaneously in the searching of qualified attributes. Accordingly, the produced computational time for obtaining reduct is high if the volume of data is great. Therefore, to decrease the elapsed time of searching multi-granularity reduct, an acceleration approach is proposed in this paper. Our proposed approach is mainly designed based on the mechanism of positive approximation, and then the selecting of qualified attributes are conducted with the gradually reduced sample space instead of the whole universe. Based on neighborhood rough set, the experimental results show us that the proposed acceleration approach can not only generate the same reduct, but also signifi-

cantly reduce the elapsed time of searching reduct compared with the general process of computing multi-granularity reduct. The following topics deserve our further research.

1. Only the measure of approximation quality is considered in this paper, our acceleration strategy will be further investigated by other measure such as information entropy.
2. The main mechanism of our proposed acceleration approach is only realized from the perspective of sample, a novel strategy, which considers the aspects of both sample and attribute, will be further investigated to improve the performance of acceleration approach.

Acknowledgements

This work was supported by the Natural Science Foundation of China (No. 61906078) and the Key Laboratory of Data Science and Intelligence Application, Fujian Province University (No. D1901).

References

[1] Hongmei Chen, Tianrui Li, Xin Fan, and Chuan Luo. Feature selection for imbalanced data based on neighborhood rough sets. *Information Sciences*, 483:1 – 20, 2019.
[2] Yan Chen, Keyu Liu, Jingjing Song, Hamido Fujita, Xibei Yang, and Yuhua Qian. Attribute group for attribute reduction. *Information Sciences*, 535:64 – 80, 2020.
[3] Xiuyi Jia, Lin Shang, Bing Zhou, and Yiyu Yao. Generalized attribute reduct in rough set theory. *Knowledge-Based Systems*, 91:204 – 218, 2016.
[4] Jingzheng Li, Xibei Yang, Xiaoning Song, Jinhai Li, Pingxin Wang, and Dong-Jun Yu. Neighborhood attribute reduction: a multi-criterion approach. *International Journal of Machine Learning and Cybernetics*, 10:731 – 742, 2019.
[5] Xiansheng Rao, Xibei Yang, Xin Yang, Xiangjian Chen, Dun Liu, and Yuhua Qian. Quickly calculating reduct: An attribute relationship based approach. *Knowledge-Based Systems*, 200:106014, 2020.
[6] Roman W. Swiniarski and Andrzej Skowron. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 24:833 – 849, 2003.
[7] Sharma Haresh Kumar, Kumari Kriti, and Kar Samarjit. A rough set approach for forecasting models. *Decision Making: Applications in Management and Engineering*, 3:1 – 21, 2020.
[8] Stankovic Milan, Gladovic Pavle, and Popovic Vladimir. Determining the importance of the criteria of traffic accessibility using fuzzy ahp and rough ahp method. *Decision Making: Applications in Management and Engineering*, 2:86 – 104, 2019.
[9] Karavidic Zoran and Projovic Damir. A multi-criteria decision-making (mcdm) model in the security forces operations based on rough sets. *Decision Making: Applications in Management and Engineering*, 1:97 – 120, 2018.
[10] Hongmei Chen, Tianrui Li, Chuan Luo, Shi-Jinn Horng, and Guoyin Wang. A decision-theoretic rough set approach for dynamic data mining. *IEEE Transactions on Fuzzy Systems*, 23:1958 – 1970, 2015.
[11] Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70 – 79, 2018.
[12] Changzhong Wang, Mingwen Shao, Qiang He, Yuhua Qian, and Yali Qi. Feature subset selection based on fuzzy neighborhood rough sets. *Knowledge-Based Systems*, 111:173 – 179, 2016.
[13] Suping Xu, Xibei Yang, Hualong Yu, Dong-Jun Yu, Jingyu Yang, and Eric C.C. Tsang. Multi-label learning with label-specific feature reduction. *Knowledge-Based Systems*, 104:52 – 61, 2016.
[14] Yuhua Qian, Jiye Liang, Deyu Li, Haiyun Zhang, and Chuangyin Dang. Measures for evaluating the decision performance of a decision table in rough set theory. *Information Sciences*, 178:181 – 202, 2008.

[15] Yuhua Qian, Jiye Liang, Witold Pedrycz, and Chuangyin Dang. Positive approximation: An accelerator for attribute reduction in rough set theory. *Artificial Intelligence*, 174:597 – 618, 2010.

[16] Zdzisław Pawlak and Andrzej Skowron. Rough sets: Some extensions. *Information Sciences*, 177:28 – 40, 2007.

[17] Xiao Zhang, Changlin Mei, Degang Chen, and Jinhai Li. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy. *Pattern Recognition*, 56:1 – 15, 2016.

[18] Yuhua Qian, Jiye Liang, and Chuangyin Dang. Knowledge structure, knowledge granulation and knowledge distance in a knowledge base. *International Journal of Approximate Reasoning*, 50:174 – 188, 2009.

[19] Wei Wei and Jiye Liang. Information fusion in rough set theory : An overview. *Information Fusion*, 48:107 – 118, 2019.

[20] Yiyu Yao. Information granulation and rough set approximation. *International Journal of Intelligent Systems*, 16:87 – 104, 2001.

[21] Witold Pedrycz and Wladyslaw Homenda. Building the fundamentals of granular computing: A principle of justifiable granularity. *Applied Soft Computing*, 13:4209 – 4218, 2013.

[22] Yuhua Qian, Honghong Cheng, Jieting Wang, Jiye Liang, Witold Pedrycz, and Chuangyin Dang. Grouping granular structures in human granulation intelligence. *Information Sciences*, 382-383:150 – 169, 2017.

[23] Qinghua Hu, Daren Yu, Jinfu Liu, and Congxin Wu. Neighborhood rough set based heterogeneous feature subset selection. *Information Sciences*, 178:3577 – 3594, 2008.

[24] Yuan Gao, Xiangjian Chen, Xibei Yang, Pingxin Wang, and Jusheng Mi. Ensemble-based neighborhood attribute reduction: A multigranularity view. *Complexity*, 2019:1 – 17, 2019.

[25] Zehua Jiang, Xibei Yang, Hualong Yu, Dun Liu, Pingxin Wang, and Yuhua Qian. Accelerator for multigranularity attribute reduction. *Knowledge-Based Systems*, 177:145 – 158, 2019.

[26] Keyu Liu, Xibei Yang, Hamido Fujita, Dun Liu, Xin Yang, and Yuhua Qian. An efficient selector for multi-granularity attribute reduction. *Information Sciences*, 505:457 – 472, 2019.

[27] Didier Dubois and Henri Prade. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17:191 – 209, 1990.

[28] Hengrong Ju, Huaxiong Li, Xibei Yang, Xianzhong Zhou, and Bing Huang. Cost-sensitive rough set: A multi-granulation approach. *Knowledge-Based Systems*, 123:137 – 153, 2017.

[29] Changzhong Wang, Yunpeng Shi, Xiaodong Fan, and Mingwen Shao. Attribute reduction based on k-nearest neighborhood rough sets. *International Journal of Approximate Reasoning*, 106:18 – 31, 2019.

[30] Xibei Yang, Shaochen Liang, Hualong Yu, Shang Gao, and Yuhua Qian. Pseudo-label neighborhood rough set: Measures and attribute reductions. *International Journal of Approximate Reasoning*, 105:112 – 129, 2019.

[31] Xibei Yang and Yiyu Yao. Ensemble selector for attribute reduction. *Applied Soft Computing*, 70:1 – 11, 2018.

[32] Dominik Slezak. Approximate entropy reducts. *Fundamenta Informaticae*, 53:365 – 390, 2002.

[33] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:1 – 27, 2011.