ACCELERATION METHOD FOR ATTRIBUTE REDUCTION BASED ON THREE-WAY DECISIONS

Doudou GUO[1], Chunmao JIANG[2]*, Anpeng LIU[3], Guanqi HE[4]

Attribution reduction is one of the key topics in the field of rough set theory. There is a redundant calculation in the traditional reduction resulting in lots of time consumption. To solve this problem, an acceleration method for attribute reduction based on sequential three-way decisions are proposed. The specific steps are as follows: 1) calculate the significance of the attribute in the decision system; 2) the attributes will be divided into three domains in terms of the significance of the corresponding attribute. And the attribute with maximal significance will be classified into the positive domain. The attributes whose significance value equal zero will be classified into the negative domain and other attributes will be classified into the boundary domain; 3) calculate the significance of the attributes in the boundary domain cyclically and divide the obtained result, until the set of attributes in the positive domain intended constraints is satisfied. In this research, 6 UCI data sets are selected to conduct experiments in the global and local reduction environments, respectively. The experimental results show that the proposed method can effectively reduce the time consumption of searching reduct in such two environments and will not contribute to a poor classification performance.

Keywords: attribute reduction; neighborhood rough set; sequential decision; significance; three-way decisions

1. Introduction

Rough set theory [1-2] is a formal tool for describing inaccurate and uncertain information. It is often used to mine internal information and refine data. Today, rough set theory has been widely used in research fields such as machine learning and knowledge discovery, data mining, decision and analysis. Attribute reduction [3-8], as the key topic in the field of rough set theory, has often attracted much attention of many researchers. The core of attribute reduction is the process of removing redundant attributes on the premise of given metrics.

[1] Master student., Dept. of Computer Science and Information Engineer, Harbin Normal University, Harbin, China, E-mail: 864782389@qq.com.
[2] Prof., Dept. of Computer Science and Information Engineer, Harbin Normal University, Harbin, China, corresponding author: hsdrose@126.com
[3] Postgraduate, Dept. of Computer Science and Information Engineer, Harbin Normal University, Harbin, China, E-mail: 120926248@qq.com
[4] Undergraduate, Shanghai University of Science and Technology, Shanghai, China, E-mail: 619117072@qq.com

With the increase of data volume and data dimension in practical applications, the redundant attributes of data also increase, which will seriously affect the computing efficiency of the computer and lead to a significant increase in the calculation time of attribute reduction. Therefore, many scholars have carried out in-depth analysis and research. Reference [9] established the reduction process of the current radius on the basis of the reduction obtained by the previous radius and reduced the calculation time of attribute reduction under multiple radii through a heuristic algorithm. Reference [10] reduces the calculation time of reduction by excluding redundant samples with the same decision value, and by using decision attributes. Reference [11] achieves the purpose of fast reduction by adding multiple attributes in each iteration of heuristic attribute reduction.

In summary, the existing acceleration methods are for the neighborhood radius, for the positive domain sample, and for adding multiple attributes at the same time. However, the problem that the redundant attributes of the samples are calculated multiple times during the iteration process is rarely studied. Based on this, an acceleration method based on the three-way decisions [12] is proposed. The proposed method considers adding attributes to the reduction set and excluding some irrelevant attributes at the same time, which aims to reduce the calculation time of attribute reduction. This paper uses the additive attribute reduction as an example to discuss the effectiveness of the three-way decisions acceleration methods, which include a global perspective and a local perspective.

2. Material and Methods

2.1. Neighborhood rough set

In rough set theory, a decision system can be described as a pair such that $DS=<U, AT\cup D>$, U is the set of all objects, called the universe; AT is the set of condition attributes; D is one set of decision attribute and $AT\cap D=\varnothing$.

Definition 1 Given a decision system DS, if the values of D are all discrete, then the division over the universe is defined as follows:

$$U/\ \mathrm{IND}(D)=\{X_1, X_2,\ldots, X_q\} \tag{1}$$

In which, for $\forall X_p \in U/\ \mathrm{IND}(D)$, X_p indicates the p-th decision class. $\mathrm{IND}(D)=\{(x, y)\in U\times U: \forall\ a_i\in AT, a_i(x)= a_i(y)\}$, $a_i(x)$ implies the the value of the sample x over the attribute a_i.

The traditional rough set model is only suitable for processing discrete data, not suitable for continuous or mixed data in practical applications. Therefore, *Hu* [13] proposed the neighborhood rough set model, which can effectively address the problem of information loss of the traditional model after data discretization. The specific method is to use the information provided by condition attributes to calculate distances and then construct neighborhood relations to process complex data.

Definition 2 [14] Given a decision system DS, given a neighborhood δ, then the neighborhood relationships can be defined as:

$$N=\{(x, y)\in U\times U : r_A(x, y)\leq \delta\} \tag{2}$$

In which $\forall x, y\in U$, $r_A(x, y)$ denotes the distance between samples x and y. Euclidean distance is employed to calculate the distance between two samples in this paper.

Then the δ neighborhood of sample x can be formulated as:

$$\delta_A(x) = \{y \mid r_A(x, y)\leq \delta\} \tag{3}$$

In which, the neighborhood radius is δ, and $\delta > 0$.

Definition 3 Given a decision system DS, $U/\mathrm{IND}(D)=\{X_1, X_2,\ldots, X_q\}$, and $\forall X_p \in U/\mathrm{IND}(D)$, the lower and upper approximation sets of X_p in terms of A can be defined as follows:

$$\underline{X_p}_A=\{x\in U \mid \delta_A(x)\subseteq X_p\} \tag{4}$$

$$\overline{X_p}_A=\{x\in U \mid \delta_A(x)\cap X_p\neq\varnothing\} \tag{5}$$

Based on the concepts of lower approximation set and upper approximation set, the universe can be divided into three parts, positive domain POS(X), boundary domain BND(X), and negative domain NEG(X). The three-way decisions based on rough set [15] is defined as follows:

$$\mathrm{POS}(X)=\underline{X_p}_A=\{x\in U \mid \delta_A(x)\subseteq X_p\} \tag{6}$$

$$\mathrm{BND}(X)=\overline{X_p}_A-\underline{X_p}_A=\{x\in U \mid \neg(\delta_A(x)\subseteq X_p)\wedge\neg(\delta_A(x)\subseteq X_P^C)\} \tag{7}$$

$$\mathrm{NEG}(X)=U-\overline{X_p}_A=\{x\in U \mid \delta_A(x)\subseteq X_p{}^C\} \tag{8}$$

It can be observed that the objects in the positive domain represent samples certainly belong to the decision class X_p, and the objects in the negative domain represent samples certainly not belong to the decision class X_p, and the objects in the boundary domain represent samples that are not sure to belong to the decision class X_p. Semantic interpretations of acceptance, rejection and delayed decision can be given separately.

2.2. Attribute reduct

There are several metrics for attribute reduction based on different applications. Approximation quality [16-17] and conditional entropy [18-20] are two common metrics. The approximation quality is employed to construct the constraint conditions of the reduction solution in this paper. The definition of the approximation quality is given as follows.

Definition 4 Given a decision system DS, $U/\mathrm{IND}(D)=\{X_1, X_2,\ldots, X_q\}$, $\forall B\subseteq AT$, the approximation quality of D with respect to B is defined as:

$$\gamma(B,D)=\frac{\left|\bigcup_{p=1}^{q}\underline{X_p}_B\right|}{|U|} \tag{9}$$

Where $|X|$ represents the number of sets X.

Obviously, $0 \le \gamma(B, D) \le 1$, $\gamma(B, D)$ indicates that according to the condition attribute B, the proportion of samples belonging to a certain decision category to the total sample is determined.

The calculation method used in this paper is a search method based on a fitness function. The significance of attributes based on approximation quality is evaluated as follows:

Definition 5 Given a decision system DS, $\forall B\subseteq AT$, B is a reduct of approximation quality in DS when it satisfies the following conditions:

(1) $\gamma(B, D)\ge\gamma(AT, D)$;

(2) $\forall B' \subseteq B$, $\gamma(B', D)<\gamma(B, D)$.

The evaluation of attribute significance is based on sample granulation. Sample granulation is the process of constructing neighborhood relations based on the information provided by conditional attributes to obtain the granulation results over the universe. According to the information granulation result, a certain metric value can be calculated. Moreover, if an attribute is added into the attribute set, then the variation of the corresponding metric value can be used to evaluate the importance of the attribute. Therefore, the attribute significance can be formulated as follows:

$$\mathrm{Sig}_\gamma(a_i,B,D)=\gamma(B\cup\{a_i\},D)-\gamma(B,D) \tag{10}$$

In which, $a_i\in AT$. In attribute reduction, the widely used search strategies include: adding attributes, deleting attributes, and adding attributes before deleting attributes. This paper uses an additive strategy to calculate the reduction, that is, the searching of attributes begins with an empty set, and then add the attributes with the most important into reduct one by one until the constraint conditions are satisfied. The traditional additive strategy to compute reduct is shown as follows:

Algorithm 1	**Traditional Additive Attribute Reduction**
Input:	Decision system DS=<U, $AT\cup D$>, neighborhood radius δ.
Output:	One approximation quality γ reduct B.
Step 1:	Compute $\gamma(AT, D)$;
Step 2:	$B\leftarrow\varnothing$;
Step 3:	**Do**
	(1) $\forall a_i\subseteq AT- B$, calculate the $Sig_\gamma(a_i, B, D)$ of attribute a_i;
	(2) b is selected as one of the most important attributes, then $B=B\cup\{b\}$;
	(3) Compute $\gamma(B, D)$;
	Until $\gamma(B, D)\ge\gamma(AT, D)$;

Step 4:	Return B.

The time complexity of Algorithm 1 is $O(|U|^2 \cdot |AT|^2)$, where $|U|$ is the sample size in the universe and $|AT|$ is the number of condition attributes. The algorithm calculates all decision categories as a whole to select the most important attributes and then obtains the reduction result. The advantage of this algorithm is that it can make effective reductions based on decision attributes, but the calculation cannot be refined into specific decision categories. In practical applications, researchers often pay more attention to those attributes that affect specific decision categories and thus develop research on ensemble attribute reduction.

Algorithm 1 is suitable for situations that are not sensitive to specific decision classes. In specific applications, ensemble reduction can balance the demand for decision attributes, which has often attracted the attention of many researchers. First, the definition of local approximation quality can be defined as follows.

Definition 6 [21] Given a decision system DS , $U/\mathrm{IND}(D)=\{X_1, X_2,\ldots, X_q\}$, $\forall B\subseteq AT$, the local approximation of X_p with respect B can be formulated as follows:

$$\gamma(B, X_p) = \frac{\left|{X_p}_B\right|}{\left|X_p\right|} \tag{11}$$

Combining formula (9) and formula (10), we can get the local attribute significance of a single decision class X_p in the ensemble environment. As is shown as follow:

$$\mathrm{Sig}_\gamma(a_i, B, X_p) = \gamma(B \cup \{a_i\}, X_p) - \gamma(B, X_p) \tag{12}$$

Based on the local attribute significance, the attributes with the maximal importance under each decision class are separately solved, and the obtained results are synthesized to obtain candidate attributes. The traditional ensemble attribute reduction algorithm is shown in Algorithm 2.

Algorithm 2	**Traditional ensemble additive reduction**
Input:	Decision system DS=<U, $AT \cup D$>, neighborhood radius δ.
Output:	One approximation quality γ reduct L.
Step 1:	Compute $\gamma(AT, D)$;
Step 2:	$L \leftarrow \varnothing$;
Step 3:	**Do**
	$T \leftarrow \varnothing$;
	For p=1:q
	(1) $\forall a_i \subseteq AT- B$, calculate the $\mathrm{Sig}_\gamma(a_i, L, X_p)$;
	(2) Select a_j, when $\mathrm{Sig}_\gamma(a_j, L, X_p)=\max\{\mathrm{Sig}_\gamma(a_i, L, X_p)$: $\forall a_i \subseteq AT- L\}$;
	(3) $T = T \cup \{a_j\}$;

	End For
	Select the attribute l that appears most frequently in T, then
	$L = L \cup \{l\}$;
	(4) Compute $\gamma(L, D)$;
	Until $\gamma(L, D) \geq \gamma(AT, D)$;
Step 4:	Return L.

The time complexity of Algorithm 2 is $O(q \cdot |U|^2 \cdot |AT|^2)$, where q is the number of decision classes, $|U|$ is the number of samples in the universe, and $|AT|$ is the number of conditional attributes. The advantage of this algorithm is that it balances the requirements of various decision attributes. The disadvantage is that the algorithm increases the amount of calculation during the iteration process, which leads to an increase in the time consumption of searching reduct.

2.3. Acceleration algorithm for attribute reduction based on three-way decisions

To reduce the time consumption of attribute reduction in an ensemble environment, an acceleration method of attribute reduction based on three-way decisions is proposed. The core of this method is the combination of sequential three-way decision ideas, excludes irrelevant attributes when selecting attributes. Reduce time consumption by reducing the number of candidate attributes. First, the construction process of sequential three-way decisions at m level is shown in Fig. 1.

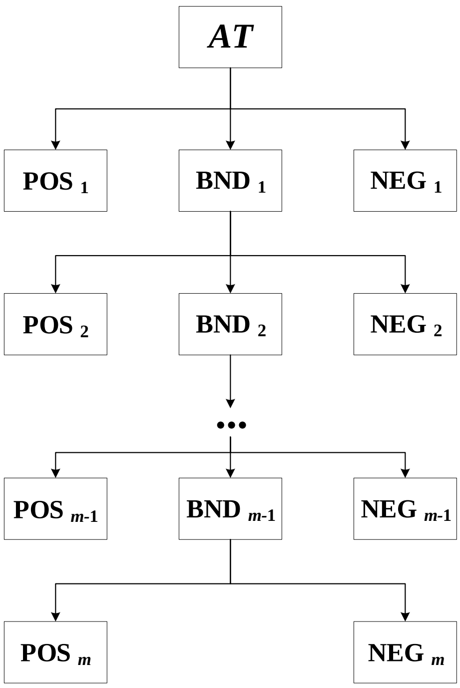


Fig.. 1. Construction process of S3WD

The candidate attributes in the BND (Delayed Decision Domain) are iterated for three parts, and finally converted into two-way decisions to obtain the reduction set B, at the same time $B=\mathrm{POS}_1 \cup \mathrm{POS}_2 \cup \ldots \cup \mathrm{POS}_m$.

Definition 7 Given the i-th level candidate attribute set $P_i(X)$, $\forall X_j \subseteq AT$, $\mathrm{Sig}(P_i(X_j))$ indicates the significance of the attribute X_j, and the rules for dividing the positive domain, the boundary domain and the negative domain of the i-th layer are as follows.

$$\mathrm{POS}_i(X) = \{x \in U_i \mid \frac{\mathrm{Sig}(P_i(X_j \mid [x]_{AT}))}{\mathrm{Sig}_{\max}(P_i(X))} = 1\} \tag{13}$$

$$\mathrm{BND}_i(X) = \{x \in U_i \mid 0 < \frac{\mathrm{Sig}(P_i(X_j \mid [x]_{AT}))}{\mathrm{Sig}_{\max}(P_i(X))} < 1\} \tag{14}$$

$$\mathrm{NEG}_i(X) = \{x \in U_i \mid \frac{\mathrm{Sig}(P_i(X_j \mid [x]_{AT}))}{\mathrm{Sig}_{\max}(P_i(X))} = 0\} \tag{15}$$

Obviously, $1 \le i \le m$, and the $\mathrm{Sig}_{\max}(P_i(X))$ represents the most important attribute of the i-th layer.

The acceleration method is applied to the attribute reduction algorithm based on the fitness function. An acceleration algorithm based on sequential three-way decisions is obtained. The algorithm selects the highest significance attribute and excludes the attribute which value equal zero. The specific steps are shown in Algorithm 3:

Algorithm 3	**Acceleration algorithm for attribute reduction**
Input:	Decision system $DS=<U, AT \cup D>$, neighborhood radius δ.
Output:	One approximation quality γ reduct B.
Step 1:	Compute $\gamma(AT, D)$;
Step 2:	$B \leftarrow \varnothing$; $\mathrm{POS} \leftarrow \varnothing$; $\mathrm{BND} \leftarrow AT$; $\mathrm{NEG} \leftarrow \varnothing$;
Step 3:	**Do**
	(1) $\forall\, a_i \subseteq AT\text{-}B\text{-}\mathrm{NEG}$, calculate the $\mathrm{Sig}_\gamma(a_i, B, D)$;
	(2) Put the most important attribute a_j into the positive domain(POS), the important value of zero are put into the negative domain(NEG), and the rest of the attributes into the boundary domain(BND);
	(3) $B=B \cup \mathrm{POS}$;
	(4) Compute $\gamma(B, D)$;
	Until $\gamma(B, D) \ge \gamma(AT, D)$;
Step 4:	Return B.

The time complexity of Algorithm 3 is $O(|U|^2 \cdot (|AT|+|\mathrm{BND}|_1+\cdots+|\mathrm{BND}|_{m-1}))$, in which $|U|$ is the number of samples in the universe, and $|\mathrm{BND}|$ is the number of attributes in the boundary domain. Compared with the traditional algorithm, the new algorithm excludes irrelevant attributes during iteration, avoiding invalid calculation of these attributes during the reduction process, and thus helps to improve the reduction efficiency.

This acceleration method is also applicable in an integrated environment. The specific steps are shown in Algorithm 4:

Algorithm 4	**Acceleration algorithm for ensemble reduction**
Input:	Decision system DS=<U, $AT \cup D$>, neighborhood radius δ.
Output:	One approximation quality γ reduct L.
Step 1:	Compute γ (AT, D);
Step 2:	$L \leftarrow \varnothing$; POS $\leftarrow \varnothing$; BND $\leftarrow AT$; NEG $\leftarrow \varnothing$;
Step 3:	**Do**
	$T \leftarrow \varnothing$;
	For p=1:q
	(1) $\forall a_i \subseteq AT$-B-NEG, calculate the $\mathrm{Sig}_\gamma (a_i, L, X_p)$;
	(2) Select a_j, when $\mathrm{Sig}_\gamma (a_j, L, X_p)$=max{ $\mathrm{Sig}_\gamma (a_i, L, X_p)$: $\forall a_i \subseteq AT$- L };
	(3) $T = T \cup \{a_j\}$;
	End For
	Select the attribute l which appears most frequently in T and put it in the positive domain(POS), the important value of zero is put into the negative domain(NEG), and the rest of the attributes into the boundary domain(BND);
	(4) $L = L \cup$ POS;
	(5) Compute $\gamma(L, D)$;
	Until $\gamma(L, D) \geq \gamma(AT, D)$;
Step 4:	Output L.

The time complexity of Algorithm 4 is $O(q \cdot|U|^2\cdot(|AT|+|\mathrm{BND}|_1+\cdots+|\mathrm{BND}|_{m-1}))$, where q is the number of decision classes, $|U|$ is the number of samples in the universe, and |BND| is the number of attributes in the boundary domain. After adding three-way decisions acceleration methods, Algorithm 4 reduces the amount of calculation compared to Algorithm 2 in the reduction calculation process.

3. Results

To demonstrate the effectiveness of the proposed acceleration method, this paper selected 6 sets of data from the UCI data set for experiments and compared the classification accuracy and time consumption of the acceleration method with the traditional method. The description of the data set is shown in Table 1. The experimental environment is a PC, dual-core 1.00GHz CPU, 4G memory, Windows 10 operating system, Matlab R2016b experimental platform.

The experiment uses the method of 5-fold cross-validation [22], and ten radii are selected, they are 0.02, 0.04, ..., 0.2.

Table 1

Data sets description

ID	Data sets	Samples	Attributes	Decision classes
1	Drug Consumption(quantified)	1885	31	7
2	Forest Type Mapping	523	28	4
3	Ionosphere Nor	351	35	2
4	Leaf	340	16	30
5	QSAR Biodegradation	1055	42	2
6	Wdbc Nor	569	31	2

The experimental data was divided into 5 parts of the same size, that is, U_1, $U_2,\ldots,U_5$. In the first round, $U_2 \cup U_3 \ldots \cup U_5$ is regarded as the training set to obtain the reduct B_1, U_1 is regarded as the testing set and uses the attributes in B_1 to derive the classification accuracy; in the second round, $U_1 \cup U_3 \ldots \cup U_5$ is regarded as the training set to obtain the reduct B_2, U_2 is regarded as the testing set and use the attributes in B_2 to derive the classification accuracy. Similarly, in the last round, $U_1 \cup U_2 \ldots \cup U_4$ is regarded as the training set to obtain the reduct B_5, U_5 is regarded as the test set, use the attributes in B_5 to derive the classification accuracy.

In the following, the approximation quality is used as a measure in the definition of attribute reduction, and the neighborhood classifier(NEC) and SVM [23] classifier are used to classify the samples in the testing set. The experiments were performed separately under the global environment and the ensemble environment. The traditional method was compared with the accelerated method using sequential three-way decisions. The time consumption and classification accuracy of these two sets of experiments were calculated and compared on the above 6 sets of data sets.

3.1. Comparison of time consumption reduction

During the experiment, the time consumption of the two algorithms to derive reducts was recorded. The detailed comparison results are shown in Table 2.

Table 2

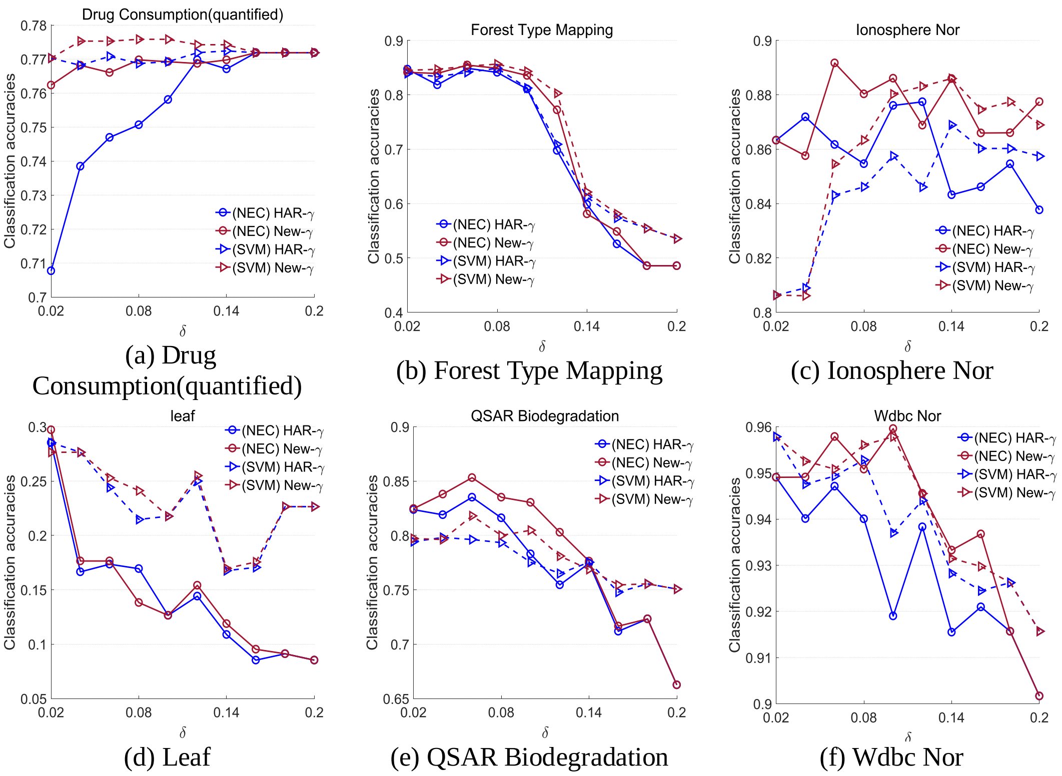
Comparison of reduction time consumption between traditional and accelerated algorithms

ID	Global reduction			Ensemble reduction		
	Traditional (s)	Acceleration (s)	Decrease (%)	Traditional (s)	Acceleration (s)	Decrease (%)
1	37.6560	11.5986	69.20	464.5959	43.7782	90.58
2	1.1740	1.1642	0.83	8.4854	4.5861	45.95
3	0.8696	0.4278	50.81	2.5411	0.7033	72.32
4	0.3439	0.2286	33.51	11.0232	1.7902	83.76
5	10.7201	9.4205	12.12	26.1192	17.7418	32.07
6	0.9668	0.8949	7.44	2.1310	1.7155	19.50

The data set names in Table 2 are replaced with the same ID numbers as in Table 1 due to space limitations. With a careful investigation of Table 2, it is not difficult to observe that the new algorithm combining the three-way decisions ideas can effectively reduce the time consumption of searching reduct, whether in the global environment or the local environment. It is worth noting that each redundant attribute is eliminated in the ensemble environment, which will reduce the calculation of reduct in every decision class at the same time. Therefore, in the same data set, the acceleration effect of the three-way decisions methods will be more obvious in the ensemble environment.

3.2. Comparison of classification accuracy of reduction results

Fig. 2 and Fig. 3, respectively, show the classification accuracy comparison between the new algorithm and the traditional algorithm in the global environment and ensemble environment. The detailed comparison results are as follows:



(a) Drug Consumption(quantified)

(b) Forest Type Mapping

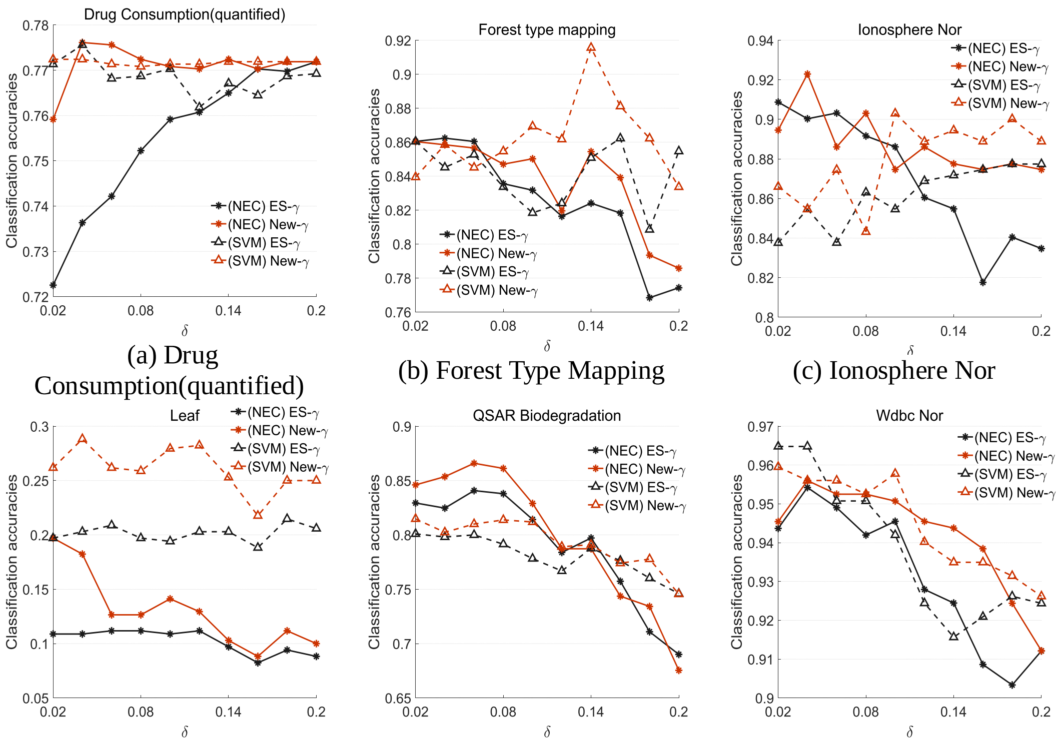
(c) Ionosphere Nor

(d) Leaf

(e) QSAR Biodegradation

(f) Wdbc Nor

Fig. 2. Comparison among classification accuracies between traditional algorithm and new algorithm in global environment



(a) Drug Consumption(quantified) (b) Forest Type Mapping (c) Ionosphere Nor

(d) Leaf (e) QSAR Biodegradation (f) Wdbc Nor

Fig. 3. Comparison among classification accuracies between traditional algorithm and new algorithm in ensemble environment

Figs. 2 show that the classification accuracy of the new method on the NEC classifier and SVM classifier in the global environment, it is not difficult to observe that the classification accuracy is not weaker than the traditional method, which can effectively guarantee the classification performance of the classifier. Figs. 3 show that the new method also has better classification performance in an ensemble environment.

4. Discussion

The acceleration method in reference [9] is to add or delete attributes to obtain the reduction set based on the attribute reduction result under the previous radius. This article considers the significance of candidate attributes in the same radius and only calculates the attributes that contribute to the significant value to achieve the purpose of accelerated reduction. Reference [11] adopts the method of adding multiple condition attributes to the reduction set at the same time to improve the calculation efficiency of reduction. In this paper, only the attributes with the highest current significance are selected each time to avoid adding irrelevant attributes to the reduction set, and the resulting reduction result will be more streamlined.

Based on the importance of attributes, this paper uses three-way decisions ideas to find the most important attributes in the set of candidate attributes and simplifies the calculation process of the next reduction by excluding attributes with an importance value of 0, thus obtaining fast attribute reduction method. This method is applied to the Global reduction and Ensemble reduction, and the experimental results of the new method can effectively reduce the calculation time and have better classification performance. In other words, the new method reduces the number of elements in the candidate attribute set, eliminates the interference of irrelevant attributes, and finally obtains better experimental results.

5. Conclusion

This paper proposes a new acceleration method for attribute reduction, which aims to solve the problem of redundant attributes being calculated multiple times in the traditional attribute reduction method. In the process of reduction calculation, we gradually eliminated redundant attributes and obtained higher reduction calculation efficiency. Through experiments, we find that the new method can significantly speed up the calculation time of reduction while ensuring the classification performance. Based on this work, the following issues deserve our further research:

(1) In this paper, only the neighborhood rough set model is used, and other rough set models such as fuzzy rough set and decision rough set will be further considered.

(2) In the process of removing attributes, the new algorithm only considers attributes with a significance of 0. Further, a dynamic threshold of attribute significance can be further set to improve the efficiency of reduction, and the cost loss of classification accuracy will be studied.

Acknowledgements

This work was supported in part by a Postgraduate Innovation Project of Harbin Normal University (HSDSSCX 2020-30).

R E F E R E N C E S

[1] *Z. Pawlak*, “Rough sets”, International journal of computer & information sciences, **vol. 11**, Oct, 1982, pp. 341–356.

[2] *G. Y. Wang, Y. Y. Yao, H. Yu*, “A survey on rough set theory and applications”, Chinese Journal of Computers, **vol. 32**, no. 7, Aug. 2009, pp. 1229-1246.

[3] *M. H. Park, H. C. Kim, S. J. Lee*, “Implementation results and service examples of GPS-Tag for indoor LBS and message service”, International conference on advanced

communication echnology (ICACT), IEEE, 2013, pp. 367- 370
[4] *J. J. Song, E. C. C. Tsang, D. G. Chen, et al*, “Minimal decision cost reduct in fuzzy decision-theoretic rough set model”, Knowledge-based systems, **vol. 126**, Jun. 2017, pp. 104-112.
[5] *H. R. Ju, H. X. Li, X. B. Yang, et al*, “Cost-sensitive rough set: a multi-granulation approach”, Knowledge-based systems, **vol. 123**, May. 2017, pp. 137-153.
[6] *J. Z. Li, X. B. Yang, H. L. Dou, et al*, “Research on ensemble significance-based attribute reduction approach”, CAAI transactions on intelligent systems, **vol. 13**, no. 3, May. 2018, pp. 414-421.
[7] *M. A. Nezhad, H. Radgohar, S. N. Salimian*, “Enhanced cultural algorithm to solve multi-objective attribute reduction based on rough set theory”, Mathematics and computers in simulation, **vol. 170**, Apr. 2020, pp. 332-350.
[8] *N. N. Thuy, S. Wongthanavasu, “*An efficient stripped cover-based accelerator for reduction of attributes in incomplete decision tables”, Expert systems with applications, **vol. 143**, Apr. 2020.
[9] *Z. H. Jiang, X. B. Yang, H. L. Yu, et al*, “Accelerator for multi-granularity attribute reduction”, Knowledge-based Systems, **vol. 177**, Aug. 2019, pp. 145-158.
[10] *N. Wang, Z. H. Peng, L. Cui*, “EasiFFRA: A fast feature reduction algorithm based on neighborhood rough set”, Journal of computer research and development, **vol. 56**, no. 12, Dec. 2019, pp. 2578-2588.
[11] *M. R. Chen, N. Zhang, X. R. Tong, et al*, “Multi-scale attribute granule based quick positive region reduction algorithm”, Journal of computer applications, vol. 39, no. 12, Dec. 2019, pp. 3426-3433.
[12] *Y. Y. Yao*, “Three-way decisions with probabilistic rough sets”, Information sciences, **vol. 180**, no. 3, Feb. 2010, 180 pp. 341-353.
[13] *Q. H. Hu, D. R. Yu, Z. X. Xie*, “Neighborhood classifiers”, Expert systems with applications, **vol. 34**, no. 2, Feb 2008, pp. 866-876.
[14] *X. B. Yang, M. Zhang, H. L. DOU*, et al, “Neighborhood systems-based rough sets in incomplete information system”, Knowledge-based systems, **vol. 24**, no. 6, Aug 2011, pp. 858-867.
[15] *D. Liu, T. R. Li, X. Yang*, et al, “Three-way decisions: research perspectives for rough sets and granular computing”, CAAI transactions on intelligent systems, **vol. 14**, no. 6, Nov 2019, pp. 1111-1120.
[16] *W. Wei, Q. Wei, F. Wang*, “Comparative study of uncertainty measure in rough set”, Journal of Nanjing University(Natural science), **vol. 51**, no. 4, May 2015, pp. 714-722.
[17] *J. S. Mi, W. Z. Wu, W. X. Zhang*, “Approaches to knowledge reduction based on variable precision rough set model”, Information sciences, **vol. 159**, no. 3, Feb 2004, pp. 255-272.
[18] *X. Zhang, C. L. Mei, D. G. Chen*, et al, “Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy”, Pattern recognition, **vol. 56**, no. 1, Aug 2016, pp. 1-15.
[19] *H. T. Liu, X. Y. Xu, J. Xie*, et al, “Incremental attribute reduction algorithm under the information view”, Journal of Chinese Computer Systems, **vol. 37**, no. 2, Feb 2016, pp. 375- 380.
[20] *H. T. Fan, T. Feng*, “The weighted conditional entropy and attribute reduction based on clustering”, Journal of Zhengzhou University(Nat. Sci. Ed), **vol. 50**, no. 1, Mar 2018, pp. 39-46.
[21] *X. B. Yang, Y. Y. Yao*, “Ensemble selector for attribute reduction”, Applied soft computing, **vol. 70**, no. 70, Sep 2018, pp. 1-11.

[22] *Dietterich, G. Thpmas*, “Approximate statistical tests for comparing supervised classification learning algorithms”, Neural computation, **vol. 10**, no. 7, Sep 1998, pp. 1895-1923.

[23] *C. C. Chih, J. L. Chih*, “LIBSVM: A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, **vol. 2**, no. 3, May 2011, pp. 1-27