

Attribute Reduction Method Based on Sequential Three-Branch Decision Model

Peiyu Su, Fu Li*

The Faculty of Mathematics and Statistics, Qinghai Nationalities University, Xining, China

Email: *fl0971@163.com

How to cite this paper: Su, P.Y. and Li, F. (2024) Attribute Reduction Method Based on Sequential Three-Branch Decision Model. *Applied Mathematics*, 15, 257-266. <https://doi.org/10.4236/am.2024.154014>

Received: December 7, 2023

Accepted: April 4, 2024

Published: April 7, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Attribute reduction is a research hotspot in rough set theory. Traditional heuristic attribute reduction methods add the most important attribute to the decision attribute set each time, resulting in multiple redundant attribute calculations, high time consumption, and low reduction efficiency. In this paper, based on the idea of sequential three-branch decision classification domain, attributes are treated as objects of three-branch division, and attributes are divided into core attributes, relatively necessary attributes, and unnecessary attributes using attribute importance and thresholds. Core attributes are added to the decision attribute set, unnecessary attributes are rejected from being added, and relatively necessary attributes are repeatedly divided until the reduction result is obtained. Experiments were conducted on 8 groups of UCI datasets, and the results show that, compared to traditional reduction methods, the method proposed in this paper can effectively reduce time consumption while ensuring classification performance.

Keywords

Attribute Reduction, Three-Branch Decision, Sequential Three-Branch Decision

1. Introduction

Rough set theory [1] is an analytical theory for handling uncertain information, and attribute reduction [2] [3], as a hotspot issue in rough set theory, aims to delete redundant attributes as much as possible while ensuring the ability of the attribute set to divide the domain remains unchanged. In recent years, research on attribute reduction has generally been divided into exhaustive methods and heuristic methods. Exhaustive methods, although able to obtain the final reduction result, are not suitable for large-scale information processing due to their

*Corresponding author.

high complexity. Heuristic methods, on the other hand, have been favored by many scholars for their high efficiency in time reduction through the use of greedy search strategies. For example, methods such as Hu [4] proposed a forward greedy attribute reduction algorithm based on attribute importance; Liu [5] improved the traditional attribute reduction algorithm proposed by Hu and others and introduced the FHARA algorithm.

Three-branch decision [6] represents a divide-and-conquer approach based on the decision rough set [7] theory, where the idea is to divide the whole into three parts and process each part separately. Yao further utilized the progressive granularity calculation concept to construct a sequential three-branch decision model [8], which could enhance processing efficiency, reduce processing costs, and is more suitable for solving complex problems. Several scholars have conducted research on the sequential three-branch decision model, for example, Ju [9] and others proposed a sequential three-branch classifier for local reduction starting from a local perspective.; Hu *et al.* used the artificial fish swarm algorithm to automatically determine the three-branch decision thresholds [10]; Sheng *et al.* proposed an attribute reduction method based on the sequential three-branch decision model [11]; Jiang *et al.* introduced an accelerated attribute reduction method utilizing the sequential three-branch decision concept [12].

The Three-Branch Decision involves dividing the domain U into positive region, negative region, and boundary region, and then proposing corresponding decision strategies. Literature [11] [12] utilizes the Three-Branch Decision concept to directly divide attributes into positive region, negative region, and boundary region. In this paper, building upon the ideas in references [11] [12], a novel attribute reduction method based on the sequential three-branch decision model is presented, where attributes are divided into core attributes, relatively necessary attributes, and unnecessary attributes using attribute importance and thresholds. This new method can exclude the influence of irrelevant attributes during the reduction process to enhance reduction efficiency and reduce time consumption. The structure of the paper is as follows: the first part reviews some basic concepts of three-branch decision and attribute reduction; the second part presents the attribute reduction method based on the sequential three-branch decision model; the third part compares this method with traditional heuristic attribute reduction algorithms; and the fourth part provides a summary.

2. Preliminary Knowledge

To better assist the research work, this section introduces some basic knowledge of three-branch decision, attribute reduction methods, and other related concepts.

2.1. Three-Branch Decision

Building upon rough set research, Yao proposed the theory of Three-Branch Decision as a common strategy for solving complex problems. In practical deci-

sion-making processes, quick judgments can be made on matters that are to be accepted or rejected. For matters that cannot be decided immediately, judgment is often delayed, leading to deferred decision-making.

Definition 1 [6]: Given a subset X of the domain U in the information system \mathcal{S} , denoted as $X \subseteq U$, the conditional probability that an object $x, x \in U$, belongs to R under the equivalence relation X in the universe U is represented as:

$$Pr(X|[x]) = \frac{|X \cap [x]_R|}{|[x]_R|} \tag{1}$$

Definition 2 [6]: the three-branch areas of the three-branch decision-making are respectively defined given the threshold value $((\alpha, \beta) | (0 \leq \beta < \alpha \leq 1))$:

$$\begin{aligned} POS(X) &= \{x \in U | Pr(X|[x]_R) \geq \alpha\} \\ BND(X) &= \{x \in U | \beta < Pr(X|[x]_R) < \alpha\} \\ NEG(X) &= \{x \in U | Pr(X|[x]_R) \leq \beta\} \end{aligned} \tag{2}$$

The calculation of the threshold (α, β) follows the Bayesian decision criterion. The calculation formula is as follows when the expected loss function is minimum and satisfies $Pr(X|[x]) + Pr(\neg X|[x]) = 1$:

$$\begin{aligned} \alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \\ \beta &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \end{aligned} \tag{3}$$

2.2. Attribute Reduction

Attribute reduction, as a core issue in rough set theory research, involves the idea of removing redundant attributes while maintaining the classification ability of an information system, in order to extract key attributes and simplify the information system. In order to represent attribute reduction based on dependency functions using information measures, Hu defined the information entropy under a decision system.

Definition 3 [4]: Given a decision system $\mathcal{S}, U = \{x_1, x_2, \dots, x_n\}, B \subseteq A$, the conditional information entropy of D with respect to B is defined as:

$$H(D|B) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|[x_i]_B \cap [x_i]_D|}{|[x_i]_B|} \tag{4}$$

where $[x_i]_B$ represents a fuzzy equivalence class, $[x_i]_D$ represents a decision class, and for $\forall a \in B, B$ is a reduction based on dependency functions if and only if $H(D|B) = H(D|A)$ and $H(D|B) > H(D|B - \{a\})$.

In order to select the decision attribute set B , the definition of attribute importance is provided, indicating the importance of each attribute to decision D .

Definition 4: Given a decision system \mathcal{S} , for $\forall a \subseteq B$, the relative attribute importance of a with respect to B is defined as:

$$Sig(a, B, D) = H(D|B) - H(D|B \cup \{a\}) \tag{5}$$

If $Sig(a, B, D) = 0$, it indicates that attribute a does not contribute to the decision, meaning that attribute a is redundant.

3. Attribute Reduction Method Based on Sequential Three-Choice Decision Model

In order to reduce the time consumption during the attribute reduction process, this paper proposes an attribute reduction method based on a sequential three-choice decision model. The core of this method lies in combining the sequential three-choice decision thinking, dividing attribute sets based on attribute importance and thresholds. For attributes in the relatively necessary set (B_{\sim}), further divisions are made. Some irrelevant attributes are excluded while selecting attributes, reducing the number of candidate attributes during iteration to decrease time consumption.

In the sequential three-choice decision model, attributes are divided into core attributes (B_+), relatively necessary attributes (B_{\sim}), and unnecessary attributes (B_-) based on attribute importance. Initially, all attributes are traversed for their importance, and then divided into B_+, B_{\sim}, B_- according to their importance. Formula 3 defines a pair of thresholds (α, β) . Attributes with an importance value $Sig(a, B, D) \geq \alpha$ are classified into B_+ , indicating their acceptance into the decision attribute set; those with $Sig(a, B, D) \leq \beta$ are classified into B_- , indicating their rejection from the decision attribute set; and attributes with $\beta < Sig(a, B, D) < \alpha$ are classified into B_{\sim} , indicating the need for further evaluation. This process is repeated until the final decision attribute set is obtained.

The acceptance, rejection, and delay decision behaviors in the three-choice decision theory correspond to the positive region, negative region, and boundary region in rough set theory, as shown in **Figure 1**. To visually observe the situation of attribute reduction, the construction process of the sequential three-choice decision model is provided in **Figure 2** (where **Figure 1** shows the division of the domain, and **Figure 2** shows the division of attributes using the sequential thinking).

Definition 5: Given a decision system S . For the i layer, a set of thresholds (α_i, β_i) is provided, and the division process is as follows:

$$\begin{aligned} B_+^i &= \{a \in B_{\sim}^{i-1} \mid Sig(a, B_{\sim}^{i-1}) \geq \alpha_i\} \\ B_{\sim}^i &= \{a \in B_{\sim}^{i-1} \mid \beta_i < Sig(a, B_{\sim}^{i-1}) < \alpha_i\} \\ B_-^i &= \{a \in B_{\sim}^{i-1} \mid Sig(a, B_{\sim}^{i-1}) \leq \beta_i\} \end{aligned} \tag{6}$$

Proposition 1: Under the conditions given in Definition 5, we have $\alpha_1 > \alpha_2 > \dots > \alpha_n$. $\beta_1 < \beta_2 < \dots < \beta_n$, that is, as i increases, α_i increases and β_i decreases, leading to an incremental increase in B_+, B_- , and a gradual decrease in B_{\sim} .

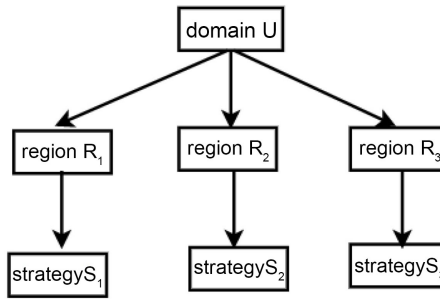


Figure 1. Three-way decision TAO model.

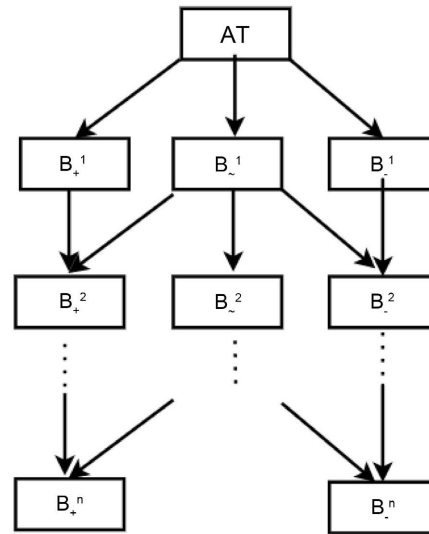


Figure 2. Sequential three-way decision model based on attribute importance.

Proof: Given a decision information system S . Based on the Bayesian decision criterion, selecting the threshold pair (α_1, β_1) that minimizes the expected loss function and satisfies the condition $Pr(X|[x]) + Pr(\neg X|[x]) = 1$, we obtain the threshold pair for GS_1 .

$$\alpha_1 = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \beta_1 = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \quad (7)$$

According to the threshold values (α_1, β_1) , the attributes in GS_1 are divided into B_+^1, B_{\sim}^1, B_-^1 , and then the threshold values (α_2, β_2) of GS_2 are calculated using formula (3), and so on. Based on formulas (1) and (3), it can be determined that $\beta_2 > \beta_1$ and $\alpha_2 < \alpha_1$. By mathematical induction, it can be concluded that $\alpha_1 > \alpha_2 > \dots > \alpha_n$ and $\beta_1 < \beta_2 < \dots < \beta_n$. Obviously, in the sequential three-way decision process, B_+ and B_{\sim} are incrementally added.

In the above process, attribute importance is used to transform B_{\sim} for many times, and threshold (α_i, β_i) is given, then the division process is as follows:

$$\begin{aligned} B_+^i &= \{a \in B_{\sim} \mid \text{Sig}(a, B_{\sim}) \geq \alpha_i\} \\ B_{\sim}^i &= \{a \in B_{\sim} \mid \beta_i < \text{Sig}(a, B_{\sim}) < \alpha_i\} \\ B_-^i &= \{a \in B_{\sim} \mid \text{Sig}(a, B_{\sim}) \leq \beta_i\} \end{aligned} \quad (8)$$

then

$$\begin{aligned}
 B_+^2 &= \{a \in B_{\sim}^1 \mid \text{Sig}(a, B_{\sim}^1) \geq \alpha_2\} \\
 B_{\sim}^2 &= \{a \in B_{\sim}^1 \mid \beta_2 < \text{Sig}(a, B_{\sim}^1) < \alpha_2\} \\
 B_-^2 &= \{a \in B_{\sim}^1 \mid \text{Sig}(a, B_{\sim}^1) \leq \beta_2\}
 \end{aligned} \tag{9}$$

...

$$\begin{aligned}
 B_+^i &= \{a \in B_{\sim}^{i-1} \mid \text{Sig}(a, B_{\sim}^{i-1}) \geq \alpha_i\} \\
 B_{\sim}^i &= \{a \in B_{\sim}^{i-1} \mid \beta_i < \text{Sig}(a, B_{\sim}^{i-1}) < \alpha_i\} \\
 B_-^i &= \{a \in B_{\sim}^{i-1} \mid \text{Sig}(a, B_{\sim}^{i-1}) \leq \beta_i\}
 \end{aligned} \tag{10}$$

In the mentioned sequential three-choice decision model, attributes in B_{\sim} are continuously divided, with useful attributes assigned to B_+ and redundant attributes assigned to B_- , until the final attribute set is obtained. The final division results are as follows:

$$\begin{aligned}
 B_+ &= B_+^1 \cup B_+^2 \cup \dots \cup B_+^n \\
 B_- &= B_-^1 \cup B_-^2 \cup \dots \cup B_-^n
 \end{aligned} \tag{11}$$

Example 1: Given a neighborhood decision information system $S = (U, AT = A \cup D)$. Where $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, $A = \{a_1, a_2, a_3, a_4\}$, $D = \{d_1, d_2, d_3\}$. According to Definition (3), assuming the values of the threshold pair (α, β) obtained based on the Bayesian criterion are $(0.8, 0.3)$. The decision information **Table 1** is shown as follows.

According to the decision attribute D , the domain U is divided into $D_1 = \{x_1, x_5, x_7, x_8\}$, $D_2 = \{x_2, x_4\}$, $D_3 = \{x_3, x_6\}$. The conditional information entropy:

$$H(D|B) = -\frac{1}{n} \sum_{i=1}^n \log \frac{\delta_B(x_i) \cap [x_i]_D}{\delta_B(x_i)} = 2.1054 \tag{12}$$

obtains $H(D|B \cup \{a_1\}) = 1.3674$, $H(D|B \cup \{a_2\}) = 2.2541$, $H(D|B \cup \{a_3\}) = 2.3625$. According to the formula $\text{Sig}(a, B, D) = H(D|B) - H(D|B \cup \{a\})$, the importance of the conditional attributes a_1, a_2, a_3, a_4 , and a_4 is obtained: $\text{Sig}(a_1, B, D) = 0.3085$, $\text{Sig}(a_2, B, D) = 0.8895$, $\text{Sig}(a_3, B, D) = 0.1721$, $\text{Sig}(a_4, B, D) = 0.8937$. Calculating the threshold pair (α_2, β_2) for GS_2 using formula (3) yields the values $(0.7, 0.3)$, indicating that attributes a_1 and a_3 can be considered redundant in the reduction process:

4. Experimental Analysis

In order to verify the effectiveness of the SAR algorithm proposed in this paper, 8 sets of UCI datasets (all downloaded from the UCI Machine Learning Database) were selected for experiments. A related comparative experiment on classification accuracy with the HAR, MSM-3WDAR, 3WDAR classification algorithms was designed. The experimental data set information is shown in **Table 2**. The

experimental environment for this paper is a Windows 10 system, with an Intel(R) Core(TM) i5-6300HQ CPU @2.30GHz computer, and the programming language used is MATLAB R2021a.

This paper utilized a 5-fold cross-validation method, selecting 10 different radii ranging from 0.02, 0.04, ..., 0.2. The experimental data was divided into 5 equal parts, namely $AT_1, AT_2, AT_3, AT_4, AT_5$. The first iteration used $AT_2 \cup AT_3 \cup AT_4 \cup AT_5$ as the training set to obtain the reduction B_+^1 , and AT_1 was used as the test set to calculate the classification accuracy using the attributes in B_+^1 . The second iteration used $AT_1 \cup AT_3 \cup AT_4 \cup AT_5$ as the training set to obtain the reduction B_+^2 , and AT_2 was used as the test set to calculate the classification accuracy using the attributes in B_+^2 . This process continued iteratively.

4.1. Comparison of Reduction Time

During the experiment, the time consumption of reduction using the 3SAR, HAR, MSM-3WDAR, and 3WDAR classification algorithms was recorded, and the comparison results are shown in [Table 3](#).

Table 1. Decision information system.

U	a_1	a_2	a_3	a_4	d
x_1	0.25	0.25	0.25	0.40	1
x_2	0.45	0.40	0.50	0.25	2
x_3	0.30	0.25	0.45	0.25	3
x_4	0.45	0.35	0.50	0.40	2
x_5	0.45	0.50	0.55	0.45	1
x_6	0.55	0.45	0.55	0.25	3
x_7	0.30	0.25	0.50	0.50	1
x_8	0.45	0.25	0.45	0.55	1

Table 2. Description of datasets.

Serial Number	Dataset Name	Samples	Conditional Attributes
1	Austra	1090	14
2	WDBC	569	30
3	Vote	432	16
4	Pima	768	8
5	QSAR	1055	41
6	Cancer	699	9
7	spectf	267	44
8	banknote	1372	4

Table 3. Comparison of reduction time among the four algorithms.

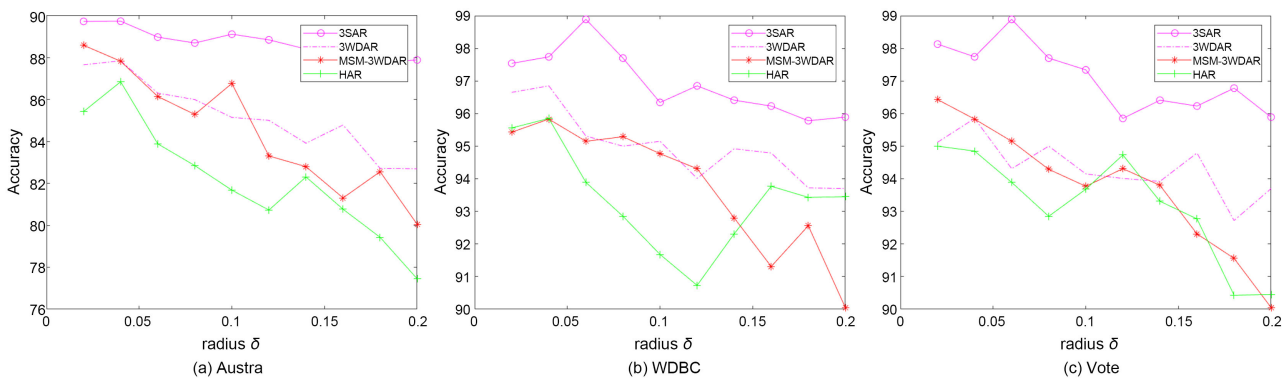
	3SAR	HAR	MSM-3WDAR	3WDAR
Austra	1.027	2.555	3.114	1.864
WDBC	1.645	3.555	3.114	1.854
Vote	0.958	5.287	2.501	4.867
Pima	1.754	4.857	3.810	2.689
QSAR	2.957	6.523	7.060	5.302
Cancer	4.274	5.571	3.851	4.058
spectf	0.351	4.388	1.582	2.539
banknote	0.967	4.577	4.610	2.059

The time complexity of 3SAR is $O(|U|^2 \cdot (|AT|^2 + |B_{\sim}^1| + \dots + |B_{\sim}^i|))$, in the iterative process, irrelevant attributes will be excluded, which will reduce the invalid calculation in the attribute reduction process, so the reduction time consumption will be reduced. The experimental results show that the new algorithm 3SAR, which integrates three decision-making ideas, consumes less reduction time than HAR, MSM-3WDAR, and 3WDAR for most radii.

4.2. Comparison of Reduction Classification Accuracy

To illustrate the effectiveness of 3SAR in classification accuracy, experiments were conducted on the samples of the 3SAR, MSM-3WDAR, and 3WDAR algorithms in the test set. The experimental results are as follows (see Figure 3).

The experimental results indicate that, for most radii, the classification accuracy obtained by 3SAR is superior to HAR, MSM-3WDAR, and 3WDAR. Furthermore, by comparing the classification accuracy of the reduction results, 3SAR demonstrates better classification performance compared to HAR, MSM-3WDAR, and 3WDAR. Additionally, 3SAR consumes less time. Therefore, based on the sequential three-decision model, the attribute reduction method effectively reduces the time consumption while ensuring the algorithm’s classification performance. This demonstrates that the new algorithm 3SAR has classification accuracy not inferior to traditional algorithms.



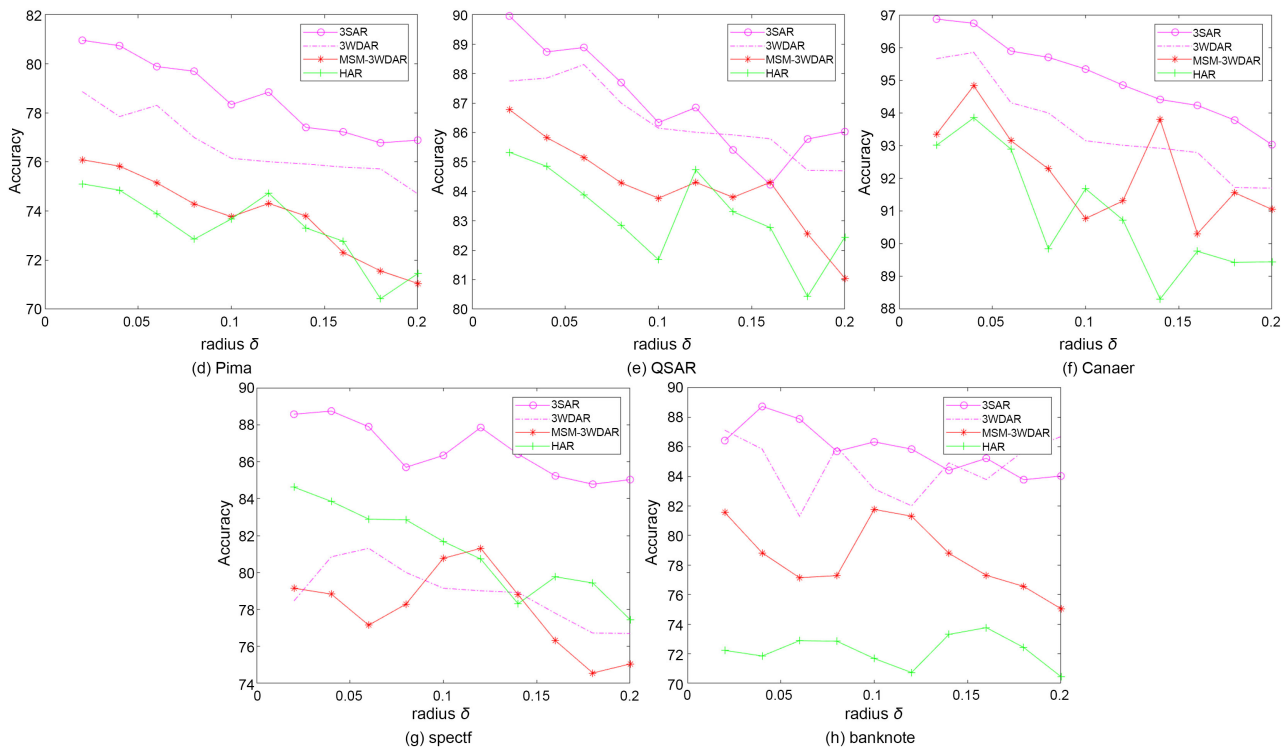


Figure 3. Comparison of accuracy among the four algorithms at different radii.

5. Conclusion

This paper proposes an attribute reduction method based on a sequential three-decision model. Building upon traditional heuristic attribute reduction methods, the importance of attributes is used as an evaluation function to determine the reduced attribute set. By utilizing attribute importance and thresholds to categorize attributes into core, relatively necessary, and unnecessary attributes, this new method is experimentally compared with HAR, MSM-3WDAR, and 3VDAR. The results show that the new algorithm improves the efficiency of reduction while maintaining reduction performance and lowering time consumption.

Funding

This work was supported by the National Natural Science Foundation of China under grant numbers 12371459.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Pawlak, Z. (1982) Rough Sets. *International Journal of Computer and Information Sciences*, **11**, 341-356. <https://doi.org/10.1007/BF01001956>
- [2] Ju, H.R., et al. (2017) Cost-Sensitive Rough Set: A Multi-Granulation Approach.

- Knowledge-Based Systems*, **123**, 137-153.
<https://doi.org/10.1016/j.knosys.2017.02.019>
- [3] Wang, G.Y., Yao, Y.Y. Yu, H. (2009) A review of Rough Set Theory and Its Applications. *Journal of Computer Research*, **32**, 1229-1246.
<https://doi.org/10.3724/SP.J.1016.2009.01229>
- [4] Hu, Q.H., Yu, D.R. Xie, Z.X. (2008) Numerical Attribute Reduction Based on Neighborhood Granulation and Rough Approximation. *Journal of Software*, **19**, 640-649. <https://doi.org/10.3724/SP.J.1001.2008.00640>
- [5] Liu, Y., *et al.* (2014) Quick Attribute Reduction Algorithm for Neighborhood Rough Set Model. *Information Sciences*, **271**, 65-81.
<https://doi.org/10.1016/j.ins.2013.08.022>
- [6] Yao, Y.Y. (2010) Three-Way Decisions with Probabilistic Rough Sets. *Information Sciences*, **180**, 341-353. <https://doi.org/10.1016/j.ins.2009.09.021>
- [7] Pawlak, Z., Wong, S.K.M. Ziarko, W. (1988) Rough Sets: Probabilistic Versus Deterministic Approach. *International Journal of Man-Machine Studies*, **29**, 81-95.
[https://doi.org/10.1016/S0020-7373\(88\)80032-4](https://doi.org/10.1016/S0020-7373(88)80032-4)
- [8] Yao, Y.Y. (2015) *Rough Sets and Three-Way Decisions*. Springer International Publishing, Berlin. https://doi.org/10.1007/978-3-319-25754-9_6
- [9] Ju, H.R., Li, H.X., Zhou, X.Z., *et al.* (2017) A Sequential Three-Decision Classifier Based on Local Attributes. *Computer Science*, **44**, 34-39+57.
- [10] Hu, P., Qin, L.X. Yao, H.M. (2016) Automatic Determination of Three-Decision Thresholds Using Artificial Fish Swarm Algorithm. *Computer Science and Modernization*, **6**, 97-102.
- [11] Sheng, R.H., Li, H.Y., Jiang, C.M., *et al.* (2021) Research on an Attribute Reduction Method Based on Sequential Three-Decision Classification. *Fuzzy Systems and Mathematics*, **35**, 48-65.
- [12] Jiang, C.M. Liu, A.P. (2020) An Accelerated Attribute Reduction Method from the Perspective of Trilateral Decisions. *Computer Engineering and Science*, **42**, 2280-2286.