Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Attribute Reduction of Neighborhood Rough Set Based on Discernment

Biqing Wang*

School of Mathematics and Computer, Tongling University, Tongling 244000, Anhui Province, China

***Corresponding author:** Biqing Wang, wbq@tlu.edu.cn

Abstract: For neighborhood rough set attribute reduction algorithms based on dependency degree, a neighborhood computation method incorporating attribute weight values and a neighborhood rough set attribute reduction algorithm using discernment as the heuristic information was proposed. The reduction algorithm comprehensively considers the dependency degree and neighborhood granulation degree of attributes, allowing for a more accurate measurement of the importance degrees of attributes. Example analyses and experimental results demonstrate the feasibility and effectiveness of the algorithm.

Keywords: Neighborhood rough set; Attribute reduction; Discernment; Algorithm

Online publication: January 23, 2024

1. Introduction

As Artificial Intelligence Generated Content (AIGC) becomes an increasingly prominent research focus, the analysis and processing of data play an even more crucial role. Rough set theory can extract essential features from a dataset while ensuring the correct classification rate [1]. The traditional Pawlak rough set model is suitable only for handling complete symbolic data. Numerical data, prevalent in real applications, cannot be processed directly; hence, data need to be discretized before reduction. However, the discretization process often results in information loss [2]. The neighborhood rough set model allows for the direct reduction of the attributes in continuous-type data, eliminating the information loss associated with discretization [3].

In the context of neighborhood rough set models, the determination of the neighborhood is of paramount importance. Typically, when calculating the neighborhood, the same weight value is assigned to all attributes. However, in practical scenarios, attributes often exhibit significant differences in the data, leading to varying degrees of importance [4]. Assigning identical weight values to different attributes neglects these differences and may result in suboptimal attribute reduction [5]. Therefore, this paper proposes to set different weight values for various conditional attributes based on experience and experiments when calculating the neighborhood. This approach addresses the challenge of the original neighborhood rough set's difficulty in distinguishing the importance of different conditional attributes.

Furthermore, existing neighborhood rough set attribute reduction algorithms measure attribute importance using dependency-based attribute importance degree as heuristic information [6]. Dependency-based attribute importance degree is defined by the difference between dependency degrees before and after the inclusion of the attribute [7]. However, this definition does not account for the neighborhood granulation degree of the attribute. In reality, the neighborhood granulation degree also reflects an attribute's ability to discriminate between objects. Two attributes causing the same dependency change cannot be considered equally important without considering their discriminative abilities through the neighborhood granulation degree. Hence, this paper introduces the concept of neighborhood granulation degree and constructs an improved evaluation metric of attribute importance, namely discernment, by combining dependency degree and neighborhood granulation degree. Based on this, a heuristic attribute reduction algorithm for neighborhood rough sets is designed, using discernment as the heuristic information. The proposed algorithm's feasibility and effectiveness are then verified through examples and experiments.

2. Preliminaries

The relevant definitions of neighborhood rough set are presented below [8,9]:

Definition 1: Given a M-dimensional real number space Ω, Δ:$R^M \times R^M \rightarrow R$, Δ is called a metric on R^M. For any y_1, y_2, $y_3 \in R^M$, if Δ meets the following conditions:

(1) $\Delta\ (y_1, y_2) \geq 0$, $y_1 = y_2$, $\Delta\ (y_1, y_2) = 0$;

(2) $\Delta\ (y_1, y_2) = \Delta\ (y_2, y_1)$;

(3) $\Delta\ (y_1, y_3) \leq \Delta\ (y_1, y_2) + \Delta\ (y_2, y_3)$;

$\langle \Omega, \Delta \rangle$ is called metric space. For any two points in space $y_i = (y_{i1}, y_{i2}, \ldots, y_{iM})$ and $y_j = (y_{j1}, y_{j2}, \ldots, y_{jM})$, the distance between y_i and y_j is often expressed in terms of the P-paradigm as:

$$\Delta_P^M(y_i, y_j) = \left(\sum_{k=1}^{M} |y_{ik} - y_{jk}|^P\right)^{1/P} \tag{1}$$

Definition 2: Given a nonempty finite set on the space of real numbers $U = (y_1, y_2, \ldots, y_n)$, for any object y_i in U, its δ neighborhood is defined as:

$\delta(y_i) = \{y | y \in U, \Delta\ (y, y_i) \leq \delta\}$(2)

where $\delta \geq 0$ is the neighborhood radius, and $\delta(y_i)$ is the neighborhood particle of y_i.

Definition 3: Let U be a set of objects, C a set of conditional attributes, and D a set of decision attributes. If C can generate a cluster of neighborhood relations on U, $NDA = \langle U, C, D \rangle$ is called a neighborhood decision-making system.

3. Proposed algorithm

In practice, the importance of various conditional attributes differs. Some attributes determine the primary characteristics of an object, while others play only a minor role in describing the object [10]. To capture this variability in conditional attributes, this paper assigns a weight value to each conditional attribute. This ensures that the calculation of distance and neighborhood reflects the primary characteristics of the object. The specific values of the weights are determined through experiments and experiences. Building upon the preceding definitions, the following definitions are provided:

Definition 4: The P-paradigm distance with attribute weight values is defined as:

$$\Delta_P^{M}{}'(y_i, y_j) = \left(\sum_{k=1}^{M} \left|\alpha_k y_{ik} - \alpha_k y_{jk}\right|^P\right)^{1/P} \tag{3}$$

where a_k is the weight of the kth attribute and $a_k \in [0, 1]$, $\sum_{k=1}^{M} a_k = 1$.

Definition 5: The δ neighborhood of y_i with attribute weight values is defined as:

$\delta'(y_i) = \{y | y \in U, \Delta_P^{M'}(y, y_i) \leq \delta\}$(4)

Existing dependency-based neighborhood rough set attribute reduction algorithms do not consider the effect of attributes on granulation degree. To solve this problem, the concept of neighborhood granulation degree is introduced below.

Definition 6: Given a neighborhood decision system $NDA = \langle U, C, D \rangle$, for any subset of attributes $Q \subseteq C$, $\delta'(y_i)$ is the neighborhood of y_i on Q. The neighborhood granulation degree of Q is defined as:

$$NGD(Q) = \frac{|\delta_Q{}'(y_1)|^2}{|U|^3} + \frac{|\delta_Q{}'(y_2)|^2}{|U|^3} + \ldots + \frac{|\delta_Q{}'(y_{|U|})|^2}{|U|^3} \tag{5}$$

Neighborhood granulation degree is a measure of the discriminative ability of an attribute set. Obviously, the smaller the radius and the more balanced the size of the neighborhood particles, the stronger the ability of the attribute set to discriminate the object, and the smaller its neighborhood granulation degree. If for all $y_i \in U$, the neighborhood of y_i on Q is $\{y_i\}$, and $NDA(Q)$ obtains the minimum value $1/|U|^2$. If for all $y_i \in U$, the neighborhood of y_i on Q is U, $NDA(Q)$ gets the maximum value of 1. In particular, $NDA(\varnothing) = 0$.

Subsequently, an improved attribute importance metric function is defined based on neighborhood granulation degree, namely discernment, was presented in the following:

Definition 7: Given a neighborhood decision system, $A \subseteq C$, for $\forall \beta \in U$, the discernment of β concerning A-β and D is defined as:

$DIS(\beta, A, D) = \Delta\gamma + \Delta NGD$(6)

where γ is the degree of dependence as defined by rough set theory [9], $\Delta\gamma = \gamma_A(D) - \gamma_{A-B}(D)$, and $\Delta NGD = NGD(A-\beta) - NGD(A)$.

From this definition, it can be seen that the improved attribute importance degree combines the dependency degree and neighborhood granulation degree. It can evaluate the classification ability of attributes more objectively, and it is a more comprehensive and precise attribute importance degree measure function. Now the improved neighborhood rough set attribute reduction algorithm is proposed using discernment as the heuristic function, as shown in **Algorithm 1**.

Algorithm 1: Neighborhood rough set attribute reduction algorithm based on discernment

Input: $NDS = \langle U, C, D \rangle$

Output: A reduction R for neighborhood decision system

$R = \varnothing$;

compute $\gamma_C(D)$;

while ($\gamma_R(D) < \gamma_C(D)$)

{for each conditional attribute $\beta \in C-R$, compute $DIS(\beta, A, D)$;

select the β with the largest discernment;

$R = R \cup \{\beta\}$;}

return R.

Obviously, the key of the algorithm is to continually select the conditional attribute with the largest discernment and add it to the reduction set until the dependency of the reduction set equals that of the original conditional attributes set. The time complexity of computing the neighborhoods for all objects is $0(|C||U|)$.

Dependency is computed based on the neighborhood of the object, resulting in a time complexity of $0(|C||U|)$. The time complexity of computing the discernment within a while loop is $0(|C|^2|U|)$. In the worst case, the discernment is computed $|C| + (|C|–1) + (|C|-2) + \ldots +3+2+1 = (|C| + 1)|C|/2$. Therefore, the time complexity of the while loop is $0(|C|^2|U|)$. Thus, the total time complexity of **Algorithm 1** is $0(|C|^4|U|)$.

4. Experimental results

The feasibility of the neighborhood rough set attribute reduction algorithm based on discernment has been verified by the example in the previous section. In this section, an experimental comparison was made between the neighborhood rough set attribute reduction algorithm based on the dependency model (Algorithm A) [11] and the proposed algorithm (Algorithm B) in this paper to further validate the effectiveness of the proposed algorithm. Three datasets – Sonar, Ionosphere, and Vehicle – from the UCI database were used in the experiments. **Table 1** presents specific information about the experimental datasets.

Table 1. Data description

Data set	Samples	$\|C\|$	Classes
Sonar	208	60	2
Ionosphere	351	34	2
Vehicle	846	18	4

The hardware and software environment for the experiment is as follows: Processor: Intel i7-7500U; Memory: 8 GB; Operating System: 64-bit Windows 10; Software: Python 3.9. Numerical attributes in the dataset often have different units of measure. This variability may affect the reduction results. For this reason, the values of all numerical conditional attributes in the dataset have been normalized to the interval [0, 1] before attribute reduction. For the selection of the neighborhood radius, the recommended value $\delta = 0.2$ was used [8]. Meanwhile, to measure the classification ability of the reduction attribute set, the classification accuracy of the reduction attribute set is calculated using an SVM classifier. The experimental results are shown in **Table 2**.

Table 2. Experimental results

Data set	Algorithm A		Algorithm B	
	$\|R\|$	Accuracy (%)	$\|R\|$	Accuracy (%)
Sonar	18	72.5	7	75.3
Ionosphere	8	89.6	6	94.1
Vehicle	10	71.0	7	76.4

As observed in **Table 2**, the number of attributes in the reduction results of the attribute reduction algorithm proposed in this paper is less than that of the neighborhood rough set attribute reduction algorithm based on the dependency model for all three datasets. Moreover, the classification accuracies of the obtained reduction results are all higher than the latter, indicating that the proposed algorithm in this paper is effective.

5. Retrospect and prospect

In this paper, a definition of discernment in conjunction with neighborhood granulation degree, building upon existing neighborhood rough set attribute reduction algorithms based on the dependency model, was provided. The basic properties of this improved attribute importance degree were discussed. Subsequently, a neighborhood rough set attribute reduction algorithm that leverages discernment's ability to discriminate between objects was proposed, and a detailed description of the algorithm's steps was provided. The feasibility of the proposed algorithm was confirmed through an illustrative example. Finally, the proposed algorithm and the traditional algorithms were compared and analyzed using three datasets from the UCI database. The experimental results demonstrated that the proposed algorithm achieved higher classification accuracy. In the next step, the algorithm could be applied to construct a knowledge graph for manufacturing production lines.

Funding

Anhui Provincial University Research Project (Project Number: 2023AH051659); Tongling University Talent Research Initiation Fund Project (Project Number: 2022tlxyrc31); Tongling University School-Level Scientific Research Project (Project Number: 2021tlxytwh05); Tongling University Horizontal Project (Project Number: 2023tlxyxdz237)

Disclosure statement

The author declares no conflict of interest.

References

[1] Pawlak Z, 1982, Rough sets. International Journal of Computer & Information Science, 11: 341–356. https://doi.org/10.1007/BF01001956

[2] Deng Z-X, Zheng Z-L, Deng D-Y, 2021, F-Neighborhood Rough Sets and Its Reduction. Acta Automatica Sinica, 47(3): 695−705. https://doi.org/10.16383/j.aas.c180556

[3] Ji J-Z, Song X-N, Yang C-C, 2020, Feature Reduction of Neighborhood Rough Set Based on Fish Swarm Algorithm in Brain Functional Connectivity. Journal of Zhejiang University (Engineering Science), 54(11): 2247–2257. https://doi.org/10.3785/j.issn.1008-973X.2020.11.020

[4] Chen T, Liu J, Zhu F, et al., 2018, A Novel Multi-Radius Neighborhood Rough Set Weighted Feature Extraction Method for Remote Sensing Image Classification. Geomatics and Information Science of Wuhan University, 43(2): 311–317. https://doi.org/10.13203/j.whugis20150290

[5] Wu Y, Zhao R, Jin W, 2019, Fault Feature Extraction of Rotating Machinery Based on EWT and a Weighted Multi Neighborhood Rough Set. Journal of Vibration and Shock, 38(24): 235–242.

[6] Wang H, Wang S, Yao J, et al., 2019, Tag Identification Rate Prediction Based on Neighborhood Rough Set and Support Vector Machine. Computer Integrated Manufacturing Systems, 25(12): 3170–3180.

[7] Zhang W, Wu W, Liang J, et al., 2001, Rough Set Theory and Method. Science Press, Beijing.

[8] Hu Q-H, Yu D-R, Xie Z-X, 2008, Numerical Attribute Reduction Based on Neighborhood Granulation and Rough Approximation. Journal of Software, 19(3): 640–649.

[9] Liu X, Wang X-L, 2021, Value Segmentation of Airline Customers Based on K-means and Neighborhood Rough Set. Operations Research and Management Science, 30(3): 104–111.

[10] Chen Y-M, Zhu Q-X, Zeng Z-Q, et al., 2018, Gene Selection Method Based on Neighborhood Rough Sets and Fish Swarm Intelligence. Journal of University of Electronic Science and Technology of China, 47(1): 99–104.

[11] Yong L, Huang W, Jiang Y, et al., 2014, Quick Attribute Reduct Algorithm for Neighborhood Rough Set Model. Information Sciences, 271: 65–81. https://doi.org/10.1016/j.ins.2014.02.093

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.