



An improved decision tree algorithm based on boundary mixed attribute dependency

Bowen Lin¹ · Caihui Liu¹ · Duoqian Miao²

Accepted: 12 December 2023 / Published online: 30 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

As an effective extension of rough set theory, the variable precision neighborhood rough set model has been applied to the attribute dependency-based improvement of decision tree algorithm of the solution concerning continuous data. However, the boundary region, as an effective description of the uncertainty of knowledge, has not been taken into account in the existing algorithms. In this paper, we define a novel decision rule based on boundary region and attribute dependency, and construct a decision tree algorithm via this decision rule. First, we define a measure called boundary coefficient based on the boundary region, which can be used for comparative quantitative analysis. Second, we define the boundary mixed attribute dependency by combining the boundary coefficient and the attribute dependency, which can consider both the boundary case of the target set and the attribute dependency. Finally, a novel decision tree algorithm is proposed by using the boundary mixed attribute dependency as the decision rule. The experimental results show that with a slight increase in leaf nodes, the total running time decreases and the maximum accuracy increases to 0.9518, which indicates the effectiveness of the proposed algorithm.

Keywords Decision tree algorithm · Neighborhood similarity · Variable precision neighborhood rough sets · Boundary coefficient · Boundary mixed attribute dependency

1 Introduction

Classification problems exist widely in people's daily lives. In the field of artificial intelligence, how to effectively tackle classification problems for humans by the application of classification algorithms has become an important research direction. Among all kinds of classification algorithms, the decision tree algorithm is considered as one of the key algorithms to solve classification problems, due to its high-quality classification results and convincing classification process. As the name suggests, the decision tree algorithm is a tree classifier based on some decision rules, and the construction

of a decision tree can be summarized into the following two parts: (1) create a split node using decision rules; (2) create a branch and a leaf node [21]. According to the above two parts, one can notice that there exist two popular research directions in decision tree learning, that is, the decision rule (also known as the partition measure) and the structure of tree classifiers. In this paper, we concentrate on the issue of decision rule. Generally speaking, the issue of decision rule exists widely in various theories. However, for a decision tree algorithm, the selection of a specific decision rule depends on whether it can construct the correct decision logic. Therefore, finding the most appropriate decision rule among various theories has become an important research direction in decision tree learning.

As well known, the representative decision tree algorithms include: Iterative Dichotomiser 3 (ID3) [24], Classifier 4.5 (C4.5) [25] and Classification And Regression Tree (CART) [1], where ID3 uses the information gain in information theory as the decision rule, C4.5 uses the information gain ratio in information theory as the decision rule, and CART uses the Gini index in probability theory as the decision rule. On the basis of the above three algorithms, many researchers attempted to construct novel decision tree

✉ Caihui Liu
liucaihui@gnnu.edu.cn; liu_caihui@163.com

Bowen Lin
lin_bw@qq.com

Duoqian Miao
dqmiao@tongji.edu.cn

¹ School of Mathematics and Computer Science, Gannan Normal University, Ganzhou 341000, Jiangxi, China

² Department of Computer Science and Technology, Tongji University, 201804 Shanghai, China

algorithms based on different theories [6]. Laber et al. [11] proposed an efficient algorithm with a penalty term in loss function to support the construction of shallow decision trees. Klaus et al. [2] proposed a gradient-based model tree algorithm which improves classification accuracy while maintaining the interpretability. In order to deal with the continuous data, the classical decision tree algorithms usually adopt the strategy of discretization. However, discretization always causes the loss of information structure, which may affect the results of classification. As a result, researchers turn to other theories with more adaptability.

Rough set theory proposed by Pawlak [23] is a widely used tool in feature selection [3, 5, 10, 20, 32, 35]. After several years of development, various variants of rough set theory have been proposed for handling complex data environments [31, 36]. For example, in order to avoid the problems caused by the rigorous equivalence relation, Ziarko proposed the variable precision rough set model, which introduced the perspective of variable precision. In order to deal with the continuous data, Lin et al. [12, 13] proposed the neighborhood rough set model, which introduced the neighborhood structure into rough sets. The neighborhood rough set model is the most commonly used method for handling continuous data in rough sets. Therefore, rough set theory, a theory with powerful generalizability, is very suitable for decision tree learning [9]. For instance, Wang et al. [27] proposed a method that combines ordinal decision trees and fuzzy rough set-based attribute reduction. Yao et al. [33] introduced the notion of attribute purity degree and proposed a decision tree algorithm based on attribute dependency and this notion. Zhai et al. [34] proposed an algorithm called “tolerance rough fuzzy decision tree”, which is based on tolerance rough fuzzy set.

As mentioned above, the neighborhood rough set model is an effective extension of Pawlak rough set model to the solution of continuous data [16, 30]. It can deal with continuous data without discretization, thus avoiding the loss of information structure caused by discretization. Plenty of scholars have explored the improved decision algorithms based on neighborhood rough set model. Xie et al. [28] proposed the neighborhood equivalence granule structure by neighborhood algebraic similarity and variable precision threshold. Moreover, they introduced the notion of neighborhood Gini index and proposed a neighborhood decision tree algorithm based on variable-precision neighborhood equivalence granules. Liu et al. [15] proposed an improved ID3 algorithm (called DIGGI) based on variable precision neighborhood rough sets. The experimental results show that the DIGGI is effective and its accuracy is greatly improved. Liu et al. [14] pointed out that there exists a neighborhood geometric structure in neighborhood system, and proposed the notion of neighborhood geometric similarity to avoid the contradiction in the transitivity of equivalence relation caused by

taking into account the neighborhood algebraic similarity only. Finally, they proposed an improved decision tree algorithm by using the attribute dependency in variable precision neighborhood rough set induced by a novel neighborhood similarity as the decision rule. However, it should be noted that the boundary region is not considered in both of the above two algorithms.

In rough set theory, the boundary region is composed of those elements which cannot be described precisely by existing knowledge when approximating a target set. Therefore, the boundary region is always considered as one of the critical clues to the uncertainty of knowledge. In fact, the boundary region can not only serve as an important explanation for the uncertainty of knowledge, but also be regarded as an essential description for the inherent extent of uncertainty of knowledge. When describing the relationship between two specific concepts in the content of the existing knowledge, the boundary region may be an appropriate option [18].

The application of the boundary region can be divided into the following two parts [19]. First, the computation of the boundary region can be used to measure whether the target set can be represented precisely by the existing knowledge, that is, the problem of whether there exists the boundary region. Second, the boundary region can also reflect the degree of difficulty that apply the existing knowledge to the concept approximation representation, that is, the problem of the size of the boundary region. Therefore, the boundary region can be used not only for the qualitative analysis of the uncertainty of knowledge, but also for the quantitative analysis.

However, there exist some problems when we attempt to apply the boundary region to the comparative quantitative analysis of multiple attributes. For instance, the computation of the boundary region involves a set of elements, which cannot be in a direct comparison. Even though the purpose of quantification can be achieved by computing the cardinality of the set of elements, since there is no limitation on the range of the boundary region, this comparison seems unconvincing. Since this comparative quantitative analysis is a novel exploration of the relationship among multiple attributes in addition to attribute dependency, it is worth studying. Furthermore, if this comparative quantitative analysis is used to describe the relationship between different condition attributes and decision attribute, we can derive the degree of difficulty when the decision attribute is approximately represented by different condition attributes respectively, that is, the influence from different condition attributes to decision respectively. Hence, the obtained boundary region-based decision rules may be suitable for decision tree learning. For instance, Luo et al. [17] proposed a new feature evaluation criterion via the combination of the dependency measure defined in the positive region and the class separability degree of the boundary region. Parthaláin et al. [22] examine a rough set-based feature selection technique that considers the num-

ber of objects in the boundary region and the distance of those objects from the lower approximation. But all of those references are the applications of specific samples in the boundary region, rather than a direct application of the boundary region itself. Therefore, in this paper, we focus on the direct application of the boundary region.

At present, there are some problems with the application of a single decision rule, even though the boundary region-based measure is used, for example, the equivalent computation results in different attributes by the single decision rule may cause the situation that the algorithm needs to select an attribute randomly, the inability to comprehensively examine attributes, and so on. Therefore, in order to avoid the equivalent computation results in different attributes, it is necessary to introduce multiple decision rules.

Hence, in order to construct the boundary region-based measure, several parts are considered. First, the boundary region-based measure should cause the same influence as the attribute dependency in mixed measure. The diverse influence may result in the preference to one of the measures in mixed measure for the attribute calculation of the result. Second, the monotonicity of both the attribute dependency and the boundary region-based measure should be the same. The diverse monotonicity may result in the mixed measure not reflecting the situation of attributes truly.

In this paper, we first use the boundary region to induce a measure that can reflect the degree of difficulty in the approximation between two attributes, called boundary coefficient. Second, we propose the notion of boundary mixed attribute dependency by combining the boundary coefficient with the attribute dependency, so that the novel measure can consider both the boundary region and the attribute dependency. Finally, we select the boundary mixed attribute dependency as the decision rule to construct a novel decision tree algorithm. The experimental results show that with a slight increase in leaf nodes, the total running time decreases and the maximum accuracy increases to 0.9518.

The contributions of this paper are given as follows. (1) We use the boundary region to induce a novel measure, i.e., boundary coefficient; (2) We propose the notion of boundary mixed attribute dependency by combining the boundary coefficient with the attribute dependency; (3) A novel decision tree algorithm is constructed based on the boundary mixed attribute dependency.

The remaining part of this paper is organized as follows. Section 2 reviews the neighborhood rough set model, the similarity in neighborhood system and the variable precision neighborhood rough set model. Section 3 gives the definitions of boundary coefficient and boundary mixed attribute dependency, and proposes the novel decision tree algorithm. Section 4 shows the experimental results. Finally, Section 5 concludes the paper.

2 Preliminaries

2.1 Neighborhood rough set model

As an extension of Pawlak rough set model, the neighborhood rough set model has received much attention from scholars, since it can deal with continuous data without discretization [13].

In rough sets, an information table is a 4-tuple $\langle U, A, f, V^{subB} \rangle$, where U is a sample set called universe, and A is an attribute set which is the description of the sample characteristics, f is an information function and V^{subB} is the value range of attribute $subB \subseteq A$. If the attribute set A is divided into two disjointed subsets C and D , then an information table is also called a decision table, where C is called the set of condition attributes and D is called the set of decision attributes.

Given a continuous information table $\langle U, A, f, V^{subB} \rangle$. For any two samples $s^a, s^b \in U$, let $subB \subseteq C$ be an attribute subset, the Minkowsky distance between s^a and s^b under $subB$ is defined as follows:

$$MDF^{subB}(s^a, s^b) = \left(\sum_{g \in subB} |s_g^a - s_g^b|^p \right)^{\frac{1}{p}}, \quad (1)$$

where s_g^a and s_g^b respectively denote the attribute values of samples s^a and s^b on attribute g , and MDF denotes the distance metric.

To make the distance calculation of samples less complex, in this paper, we set $p = 1$, and in this case MDF is also called the Manhattan distance.

Given a continuous information table $\langle U, A, f, V^{subB} \rangle$. For any sample $s^a \in U$ and attribute subsets $subB \subseteq C$, let η denote the neighborhood radius, the η -neighborhood class of s^a under the attribute subset $subB$ is defined as follows [7]:

$$N_\eta^{subB}(s^a) = \{s^b \mid MDF^{subB}(s^a, s^b) \leq \eta\}. \quad (2)$$

Formula (2) is also called the η -neighborhood granulation, and $N_\eta^{subB}(s^a)$ is also called the η -neighborhood granule of s^a .

Let $\langle U, NR \rangle$ be a neighborhood approximation space, where U represents the universe and NR represents the neighborhood relation, for any target set $X \subseteq U$, the lower and the upper approximation of the target set X can be defined as follows [7]:

$$\underline{NR}X = \{s^a \in U \mid \eta(s^a) \subseteq X\}. \quad (3)$$

$$\overline{NR}X = \{s^a \in U \mid \eta(s^a) \cap X \neq \emptyset\}. \quad (4)$$

2.2 Similarity in neighborhood system

In [14], Liu et al. introduced four similarity measures into neighborhood system and constructed four neighborhood algebraic similarities. These four neighborhood algebraic similarities can reflect the degree of similarity between two neighborhood granules. Moreover, Liu et al. pointed out that the neighborhood system has spatial properties. Hence, in terms of the distance between two neighborhood granules, they defined the geometric similarity. By combining the two kinds of similarities mentioned above, four novel neighborhood similarities was constructed.

2.2.1 Neighborhood similarity

Neighborhood similarity is a similarity constructed by combining neighborhood algebraic similarity and neighborhood geometric similarity. It considers both algebraic and geometric perspectives in neighborhood system and can tackle the transitivity contradiction of the equivalence relation.

Definition 1 [14] For any $s^a, s^b \in U$ and attribute subset $subB \subseteq C$, let $MDF^{subB}(s^a, s^b)$ represent the Minkowsky distance of sample s^a and sample s^b under attribute subset $subB$, and let η represent the neighborhood radius. Four neighborhood similarities are respectively defined as follows.

1. The neighborhood Jaccard geometric similarity under attribute subset $subB$ is defined as:

$$NRJGS^{subB}(s^a, s^b) = \left(1 - \frac{MDF^{subB}(s^a, s^b)}{2\eta}\right) \times \frac{|N_\eta^{subB}(s^a) \cap N_\eta^{subB}(s^b)|}{|N_\eta^{subB}(s^a) \cup N_\eta^{subB}(s^b)|}. \quad (5)$$

2. The neighborhood Optimistic geometric similarity under attribute subset $subB$ is defined as:

$$NROGS^{subB}(s^a, s^b) = \left(1 - \frac{MDF^{subB}(s^a, s^b)}{2\eta}\right) \times \frac{|N_\eta^{subB}(s^a) \cap N_\eta^{subB}(s^b)|}{\min(|N_\eta^{subB}(s^a)|, |N_\eta^{subB}(s^b)|)}. \quad (6)$$

3. The neighborhood Pessimistic geometric similarity under attribute subset $subB$ is defined as:

$$NRPGS^{subB}(s^a, s^b) = \left(1 - \frac{MDF^{subB}(s^a, s^b)}{2\eta}\right) \times \frac{|N_\eta^{subB}(s^a) \cap N_\eta^{subB}(s^b)|}{\max(|N_\eta^{subB}(s^a)|, |N_\eta^{subB}(s^b)|)}. \quad (7)$$

4. The neighborhood Average geometric similarity under attribute subset $subB$ is defined as:

$$NRAGS^{subB}(s^a, s^b) = \left(1 - \frac{MDF^{subB}(s^a, s^b)}{2\eta}\right) \times \frac{|N_\eta^{subB}(s^a) \cap N_\eta^{subB}(s^b)|}{\text{average}(|N_\eta^{subB}(s^a)|, |N_\eta^{subB}(s^b)|)}. \quad (8)$$

In this paper, we use $NRS^{subB}(s^a, s^b)$ to represent the above four neighborhood similarities.

2.3 Variable precision neighborhood rough set model induced by neighborhood similarity

Based on the neighborhood similarity proposed in Section 2.2.1, we can construct a variable precision neighborhood rough set model. In this model, the classical neighborhood equivalence relation is replaced by the variable precision neighborhood equivalence relation induced by neighborhood similarity, which can avoid the problems caused by the rigorous neighborhood equivalence relation.

Due to the addition of variable precision threshold, the variable precision neighborhood equivalence relation can be constructed. In order to make difference, in this paper, the decision table $\langle U, A, f, V^{subB} \rangle$ is replaced by the variable neighborhood decision table $\langle U, A, f, V^{subB}, \eta, \mu \rangle$.

Definition 2 [14] Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$ and any tuple $(s^a, s^b) \in U \times U$, the variable precision neighborhood equivalence relation is defined as follows:

$$VPNER_{(\eta, \mu)}^{subB} = \{(s^a, s^b) \in U \times U \mid NRS^{subB}(s^a, s^b) \geq \mu\}. \quad (9)$$

Definition 3 [14] Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any sample $s \in U$ any attribute subset $subB \subseteq C$, the variable precision neighborhood equivalence class of sample s induced by variable precision neighborhood equivalence relation is defined as follows:

$$U / VPNER_{(\eta, \mu)}^{subB} = \left\{ [s]_{VPNER_{(\eta, \mu)}^{subB}} \mid s \in U \right\} = \left\{ S_{(\eta, \mu)}^1, \dots, S_{(\eta, \mu)}^X \right\}. \quad (10)$$

where $S_{(\eta, \mu)}^X$ represents the variable precision neighborhood equivalence class set formed.

When $\mu = 1$, the variable precision neighborhood granule degenerates into the rigorous equivalent neighborhood granule. When $\eta = 0$ and $\mu = 1$, the variable precision neighborhood granule degenerates into the Pawlak equivalent granule. So the variable precision neighborhood granules are actually a specific type of neighborhood granules.

Definition 4 [14] Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. Assume that $U/D = \{Y^1, Y^2, \dots, Y^Z\}$ denotes the partition of U induced by D , for any attribute subset $subB \subseteq C$, the lower approximation of D and the upper approximation of D are respectively defined as follows:

$$\underline{VPNER}_{(\eta, \mu)}^{subB} D = \bigcup_{z=1}^Z \underline{VPNER}_{(\eta, \mu)}^{subB} Y^z, \tag{11}$$

$$\overline{VPNER}_{(\eta, \mu)}^{subB} D = \bigcup_{z=1}^Z \overline{VPNER}_{(\eta, \mu)}^{subB} Y^z, \tag{12}$$

where

$$\underline{VPNER}_{(\eta, \mu)}^{subB} Y^z = \bigcup \left\{ S_{(\eta, \mu)}^x \mid S_{(\eta, \mu)}^x \subseteq Y^z, S_{(\eta, \mu)}^x \in U / \underline{VPNER}_{(\delta, \beta)}^{subB} \right\},$$

$x = 1, 2, \dots, X; z = 1, 2, \dots, Z.$

$$\overline{VPNER}_{(\eta, \mu)}^{subB} Y^z = \bigcup \left\{ S_{(\eta, \mu)}^x \mid S_{(\eta, \mu)}^x \in U / \overline{VPNER}_{(\eta, \mu)}^{subB} : S_{(\eta, \mu)}^x \cap Y^z \neq \emptyset \right\},$$

$x = 1, 2, \dots, X; z = 1, 2, \dots, Z.$

In the above formula, $\underline{VPNER}_{(\eta, \mu)}^{subB} D$ is called the positive region of D , denoted as $POS_{(\eta, \mu)}^{subB} D$.

Definition 5 [14] Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$, the attribute dependency of D on $subB$ is defined as follows:

$$r_{(\eta, \mu)}(D, subB) = \frac{|\underline{VPNER}_{(\eta, \mu)}^{subB} D|}{|U|}. \tag{13}$$

3 An improved decision tree algorithm based on boundary mixed attribute dependency

In this section, we introduce the boundary coefficient, and propose a novel decision rule — boundary mixed attribute dependency, which is induced by the boundary coefficient and attribute dependency. Moreover, an improved decision tree algorithm is constructed by the novel decision rule.

3.1 Boundary coefficient and boundary mixed attribute dependency

In rough set theory, the boundary region is used to describe the uncertainty when approximating a target set using the existing knowledge. The larger the uncertainty, the larger the boundary region. Therefore, the boundary region is a good perspective for constructing the decision rule in decision tree learning. In this subsection, we first propose a novel measure called boundary coefficient, and then propose the notion of boundary mixed attribute dependency by combining the boundary coefficient with the attribute dependency.

Definition 6 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$, the boundary region of D with respect to $subB$ is defined as follows:

$$BND_{(\eta, \mu)}^{subB}(D) = \overline{VPNER}_{(\eta, \mu)}^{subB} D - \underline{VPNER}_{(\eta, \mu)}^{subB} D. \tag{14}$$

Definition 7 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$, the boundary coefficient of D with respect to $subB$ is defined as follows:

$$BC_{(\eta, \mu)}^{subB}(D) = 1 - \frac{|BND_{(\eta, \mu)}^{subB}(D)|}{|U|}. \tag{15}$$

The basic idea of the definition of boundary coefficient can be divided into two parts. (1) To ensure that the boundary coefficient and the attribute dependency can cause the same weighting influence on the final mixed decision rule, the value range of boundary coefficient should be $[0, 1]$. So we use the cardinality of the universe to normalize the cardinality of the boundary region, that is, $\frac{|BND_{(\eta, \mu)}^{subB}(D)|}{|U|}$. Therefore, the computation of the final mixed decision rule is not biased towards the boundary coefficient or the attribute dependency.

(2) To ensure that the performance of boundary coefficient to the final mixed decision rule is the same as that of attribute dependency, the monotonicity of boundary coefficient should be the same as the attribute dependency. That is, the larger the attribute dependency, the larger the boundary coefficient.

Since the monotonicity of $\frac{|BND_{(\eta, \mu)}^{subB}(D)|}{|U|}$ is opposite to the attribute dependency, we use $1 - \frac{|BND_{(\eta, \mu)}^{subB}(D)|}{|U|}$ as the boundary coefficient of D with respect to $subB$. Therefore, we can avoid the situation that the larger the attribute dependency, the smaller the boundary coefficient, so that the computation result of the final mixed decision rule can correctly reflect the true situation of attributes.

The boundary coefficient is a measure induced by the boundary region, which can be used to reflect the roughness degree or the difficulty degree during the approximation to the target set.

Property 1 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$, we have that:

- (1) $BC_{(\eta, \mu)}^{subB}(D) \in [0, 1]$;
- (2) If $BC_{(\eta, \mu)}^{subB}(D) = 0$, then $r_{(\eta, \mu)}(D, subB) = 0$;
- (3) If $BC_{(\eta, \mu)}^{subB}(D) = 1$ and $\overline{VPNER_{(\eta, \mu)}^{subB}}D = U$, then $r_{(\eta, \mu)}(D, subB) = 1$.

Proof (1) When $BC_{(\eta, \mu)}^{subB}(D)$ takes the minimum value, it means that $|BND_{(\eta, \mu)}^{subB}(D)|$ takes the maximum value, i.e., $|U|$. So we can obtain that the minimum value of $BC_{(\eta, \mu)}^{subB}(D)$ is 0. When $BC_{(\eta, \mu)}^{subB}(D)$ takes the maximum value, it means that $|BND_{(\eta, \mu)}^{subB}(D)|$ takes the minimum value, i.e., 0. So we can obtain that the maximum value of $BC_{(\eta, \mu)}^{subB}(D)$ is 1.

- (2) If $BC_{(\eta, \mu)}^{subB}(D) = 0$, then we have that $\overline{VPNER_{(\eta, \mu)}^{subB}}D - VPNER_{(\eta, \mu)}^{subB}D = U$, which means that $\overline{VPNER_{(\eta, \mu)}^{subB}}D = U$ and $VPNER_{(\eta, \mu)}^{subB}D = \emptyset$. Hence, we have that $r_{(\eta, \mu)}(D, subB) = 0$.
- (3) If $BC_{(\eta, \mu)}^{subB}(D) = 1$ and $\overline{VPNER_{(\eta, \mu)}^{subB}}D = U$, then we can obtain that $\overline{VPNER_{(\eta, \mu)}^{subB}}D - VPNER_{(\eta, \mu)}^{subB}D = \emptyset$, which means that $\overline{VPNER_{(\eta, \mu)}^{subB}}D = VPNER_{(\eta, \mu)}^{subB}D = U$. Hence, we have that $r_{(\eta, \mu)}(D, subB) = 1$.

Property 1 discusses the value range of $BC_{(\eta, \mu)}^{subB}(D)$. From Property 1, we can clearly notice that the value range of $BC_{(\eta, \mu)}^{subB}(D)$ is $[0, 1]$, which ensures that the final mixed decision rule can obtain the same impact from the boundary coefficient and the attribute dependency. Moreover, (2) and (3) of Property 1 ensure that when the boundary coefficient takes the maximum value and the minimum value, the value of attribute dependency is the same. \square

Theorem 1 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any $subB^1, subB^2 \subseteq C$, if $subB^1 \subseteq subB^2$, then we have that $BC_{(\eta, \mu)}^{subB^1}(D) \leq BC_{(\eta, \mu)}^{subB^2}(D)$.

Proof Since $subB^1 \subseteq subB^2$, we can obtain that the partition of $subB^1$ is rougher than that of $subB^2$, that is,

for any $s \in U$, $[s]_{VPNER_{(\eta, \mu)}^{subB^2}} \subseteq [s]_{VPNER_{(\eta, \mu)}^{subB^1}}$. Hence, we have that $POS_{(\eta, \mu)}^{subB^1}D \subseteq POS_{(\eta, \mu)}^{subB^2}D$, which means that $\frac{|BND_{(\eta, \mu)}^{subB^2}(D)|}{|U|} \leq \frac{|BND_{(\eta, \mu)}^{subB^1}(D)|}{|U|}$. So we can obtain that $BC_{(\eta, \mu)}^{subB^1}(D) \leq BC_{(\eta, \mu)}^{subB^2}(D)$.

From Theorem 1, it can be seen that the more knowledge there is, this is, $subB^1 \subseteq subB^2$, the deeper realization to the target set can be achieved, this is, $[s]_{VPNER_{(\eta, \mu)}^{subB^2}} \subseteq [s]_{VPNER_{(\eta, \mu)}^{subB^1}}$. Hence, we can obtain more information to make a decision, that is, the boundary coefficient will be larger. So the change of the boundary coefficient is consistent with the way we learn the rules of the world. \square

Definition 8 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$, by combining the boundary coefficient with the attribute dependency, the boundary mixed attribute dependency of D with respect to $subB$ is defined as follows:

$$BMAD_{(\eta, \mu)}^{subB}(D) = BC_{(\eta, \mu)}^{subB}(D) * r_{(\eta, \mu)}(D, subB) = (1 - \frac{|BND_{(\eta, \mu)}^{subB}(D)|}{|U|}) * \frac{|VPNER_{(\eta, \mu)}^{subB}D|}{|U|}. \tag{16}$$

Since the boundary mixed attribute dependency is constructed by combining the boundary coefficient and the attribute dependency, it can reflect both the uncertainty and the dependency of D with respect to $subB$. Therefore, the boundary mixed attribute dependency is a good decision rule for decision tree learning.

Property 2 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any attribute subset $subB \subseteq C$, we have that:

- (1) $BMAD_{(\eta, \mu)}^{subB}(D) \in [0, 1]$;
- (2) If $BMAD_{(\eta, \mu)}^{subB}(D) = 0$, then $r_{(\eta, \mu)}(D, subB) = 0$ or $BC_{(\eta, \mu)}^{subB}(D) = 0$;
- (3) If $BMAD_{(\eta, \mu)}^{subB}(D) = 1$, then $r_{(\eta, \mu)}(D, subB) = 1$ and $BC_{(\eta, \mu)}^{subB}(D) = 1$.

Proof According to Property 1, it is easy to prove Property 2, so we omit the details. \square

Table 1 A variable neighborhood decision table

U	att ₁	att ₂	d
s ¹	0.4	0.3	2
s ²	0.9	0.4	2
s ³	1	0.6	2
s ⁴	0.5	0.2	1

Theorem 2 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table, where η represents the neighborhood radius and μ represents the variable precision threshold. For any $subB^1, subB^2 \subseteq C$, if $subB^1 \subseteq subB^2$, then we have that $BMA D_{(\eta, \mu)}^{subB^1}(D) \leq BMA D_{(\eta, \mu)}^{subB^2}(D)$.

Proof Since $subB^1 \subseteq subB^2$, according to Theorem 1, we can obtain that $POS_{(\eta, \mu)}^{subB^1} D \subseteq POS_{(\eta, \mu)}^{subB^2} D$. By Definition 5, we can further obtain that $r_{(\eta, \mu)}(D, subB^1) \leq r_{(\eta, \mu)}(D, subB^2)$. Moreover, according to Theorem 1, we can obtain that $BC_{(\eta, \mu)}^{subB^1}(D) \leq BC_{(\eta, \mu)}^{subB^2}(D)$. Therefore, by Definition 8, we have that $BMA D_{(\eta, \mu)}^{subB^1}(D) \leq BMA D_{(\eta, \mu)}^{subB^2}(D)$.

Theorem 2 discusses the monotonicity of $BMA D_{(\eta, \mu)}^{subB}(D)$. From Theorem 2 we can notice that the more knowledge we have, the larger the attribute dependency, and the larger $BMA D_{(\eta, \mu)}^{subB}(D)$. Hence, the change of the boundary mixed attribute dependency is also consistent with the way we learn the rules of the world.

Example 1 Let $\langle U, A, f, V^{subB}, \eta, \mu \rangle$ be a variable neighborhood decision table (as shown in Table 1), where U represents the universe, $\{att_1, att_2\}$ represents the set of condition attributes, and $\{d\}$ represents the set of decision attributes. Suppose that the neighborhood radius η equals 0.2 and the variable precision threshold μ equals 0.7 (Tables 2, 3, 4, and 5).

From Table 6, we can obtain the following conclusions. On the one hand, the boundary mixed attribute dependency of D with respect to att_1 is larger than that of D with respect to att_2 . On the other hand, the evaluation effect of the boundary coefficient on the two condition attributes is similar to the attribute dependency. Hence, Example 1 shows the effectiveness of the boundary coefficient and the boundary mixed attribute dependency.

Table 2 The results of neighborhood granulation

U	$\eta^{att_1}(s)$	$\eta^{att_2}(s)$
s ¹	{s ¹ , s ² }	{s ¹ , s ² , s ⁴ }
s ²	{s ¹ , s ² , s ³ , s ⁴ }	{s ¹ , s ² , s ³ }
s ³	{s ² , s ³ }	{s ² , s ³ }
s ⁴	{s ¹ , s ² , s ⁴ }	{s ¹ , s ² , s ⁴ }

Table 3 The results of neighborhood similarity on att₁

samples	neighborhood Jaccard similarity	neighborhood geometric similarity	neighborhood similarity
(1, 2)	0.5	0.5	0.25
(1, 3)	0.333	0	0
(1, 4)	0.666	0.75	0.4995
(2, 3)	0.5	0.5	0.25
(2, 4)	0.75	0.75	0.5625
(3, 4)	0.25	0.25	0.0625

3.2 Algorithm design

In this subsection, a novel decision tree algorithm is constructed by regarding the boundary mixed attribute dependency as the partition measure.

Before constructing the algorithm, it should be noted that the value of neighborhood radius plays an important role in neighborhood granulation. However, during the most applications of neighborhood rough set model, the value of neighborhood radius is given by the authors subjectively. Therefore, in this paper, the improved neighborhood radius in [14] is used.

Definition 9 [14] Let $subB \subseteq C$ be a subset of condition attributes, $sta(subB)$ represent the standard deviation of $subB$, NRC represent the neighborhood radius coefficient, and \bar{s} represent the average value of $subB$, a novel neighborhood radius is defined as follows:

$$\eta^{subB} = \frac{sta(subB)}{NRC}, \tag{17}$$

where $sta(subB) = \sqrt{\frac{1}{G} \sum_{g=1}^G (s^g - \bar{s})^2}$ and s^g denote the attribute values of samples s on attribute g .

The neighborhood radius coefficient is a constant used to control the size of neighborhood radius. In this paper, we set $NRC = 2$. And the flowchart is shown as Fig. 1

Table 4 The results of neighborhood similarity on att₂

samples	neighborhood Jaccard similarity	neighborhood geometric similarity	neighborhood similarity
(1, 2)	0.5	0.75	0.375
(1, 3)	0.25	0.25	0.0625
(1, 4)	1	0.75	0.75
(2, 3)	0.666	0.5	0.333
(2, 4)	0.5	0.5	0.25
(3, 4)	0.25	0	0

Table 5 The results of neighborhood equivalence relation

U	att ₁	att ₂
s ¹	{s ¹ }	{s ¹ , s ⁴ }
s ²	{s ² }	{s ² }
s ³	{s ³ }	{s ³ }
s ⁴	{s ⁴ }	{s ¹ , s ⁴ }

The time complexity of Algorithm 1 is mainly related to Step 1. During the neighborhood granulation, all samples need to calculate the neighborhood classes of them and the time complexity is $O(|C| * |U|^2)$. The subsequent calculation of equivalence partition has the same time complexity as the neighborhood granulation, which requires a variable precision equivalent over all samples, and the time complexity is $O(|C| * |U|^2)$. Moreover, the calculation of the partition measure is a computation operation for each attribute, so the time complexity is $O(|C|)$. Therefore, the time complexity of Algorithm 1 is $O(|C| * |U|^2)$.

The time complexities of the others algorithms are conducted carefully in literature [15]. So in this paper, the details are omitted. The time complexity of algorithm in [14] is $O(|C| * |U|^2)$. The time complexity of ID3 algorithm is $O(|C| * |U| * |logU|)$. The time complexity of C4.5 algorithm is $O(|C| * |U| * |logU|)$. The time complexity of CART algorithm is $O(|C| * |U| * |logU|)$.

4 Experimental analysis

In this section, we use simulation experiments to verify the effectiveness of Algorithm 1. We choose 14 datasets from the UCI Machine Learning Repository. Table 7 details the information of the 14 datasets. We select the ID3 [24], C4.5 [25], CART [1] algorithms (for these three algorithms, the continuous attributes in each dataset are discretized into 5 intervals by the equal distance partition) and the algorithm proposed in [14] to compare with Algorithm 1. Moreover, we use three evaluation metrics: accuracy, number of leaves, and

Table 6 The results of computation

item	att ₁	att ₂
lower approximation	{s ¹ , s ² , s ³ , s ⁴ }	{s ¹ , s ² , s ³ , s ⁴ }
upper approximation	{s ¹ , s ² , s ³ , s ⁴ }	{s ² , s ³ }
boundary region	∅	{s ¹ , s ⁴ }
attribute dependence	1	0.5
boundary coefficient	1	0.5
boundary mixed attribute dependency	1	0.25

Algorithm 1 Decision tree algorithm based on the boundary mixed attribute dependency.

Input: a variable neighborhood decision table $VNRDT = \langle U, C \cup D, \eta, \mu \rangle$

Output: A variable precision neighborhood decision tree

1. for each $subB \subseteq C$ do

 Compute $BMAD_{(\eta, \mu)}^{subB}(D) = (1 - \frac{|BND_{(\eta, \mu)}^{subB}(D)|}{|U|}) * \frac{|VPNER_{(\eta, \mu)}^{subB}D|}{|U|}$;

 end for

2. Select an attribute subset $subB' \subset C$ with the maximum boundary mixed attribute dependency, that is, $subB' = arg \max_{subB \subseteq C} BMAD_{(\eta, \mu)}^{subB}(D)$, and randomly select an optimal attribute in $subB'$ as the splitting node;

3. Obtain the variable precision neighborhood equivalent granules $U/VPNER_{(\eta, \mu)}^{subB'} = \{S_{subB'}^1, S_{subB'}^2, \dots, S_{subB'}^x\}$ constructed by the variable precision neighborhood equivalence relation by the selected attribute $subB'$, and then divide the selected attribute into x variable precision neighborhood decision information subtables, and each granule is a subtable and a branch;

 For each variable precision neighborhood decision information subtable do

 4. If all decision labels in certain one variable precision neighborhood equivalent granule are the same, then a leaf node is created and named as the decision label;

 5. If $subB' = \emptyset$, then a leaf node is created and named as the decision label with the maximum number of all decision labels in the parent node;

 6. If $subB' \neq \emptyset$ and there exist different decision labels in certain one variable precision neighborhood equivalent granule, then go to step 1.

 End For

7. Return the variable precision neighborhood decision tree.

running time to assess the effectiveness of each algorithm. The experiments are repeated 4 times. In each iteration, the 10-fold cross-validation method is used to verify the accuracy and the number of leaves. Finally, we compute the average values of accuracy and number of leaves. Moreover, the running time is the total time of the 4 iterations. Since the ID3, C4.5 and CART algorithms use the equal distance partition to discretize the continuous attributes in each dataset, the running time of them is very short. Hence, in terms of the running time, we only compare the algorithm proposed in [14] with Algorithm 1.

In the experiments, the variable precision threshold μ is set to 0.8. The experiments are run in MATLAB2021b, and the hardware environment of experiments is Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz and 16.0 GB RAM.

Since there are four similarity measures for both of literature [14] and the Algorithm 1, which results in four diverse experiment results. To make the comparisons more clear, in the remainder of this paper, we use $AD(NRJGS)$ to represent the decision tree algorithm proposed in [14] which is based on $NGS * Jaccard$, and use $BMAD(NRJGS)$ to represent Algorithm 1 which is based on $NRJGS$. The other representations are the same. In addition, we do comparisons on the algorithms with the same neighborhood algebraic similarity. For instance, the $AD(NRJGS)$ is com-

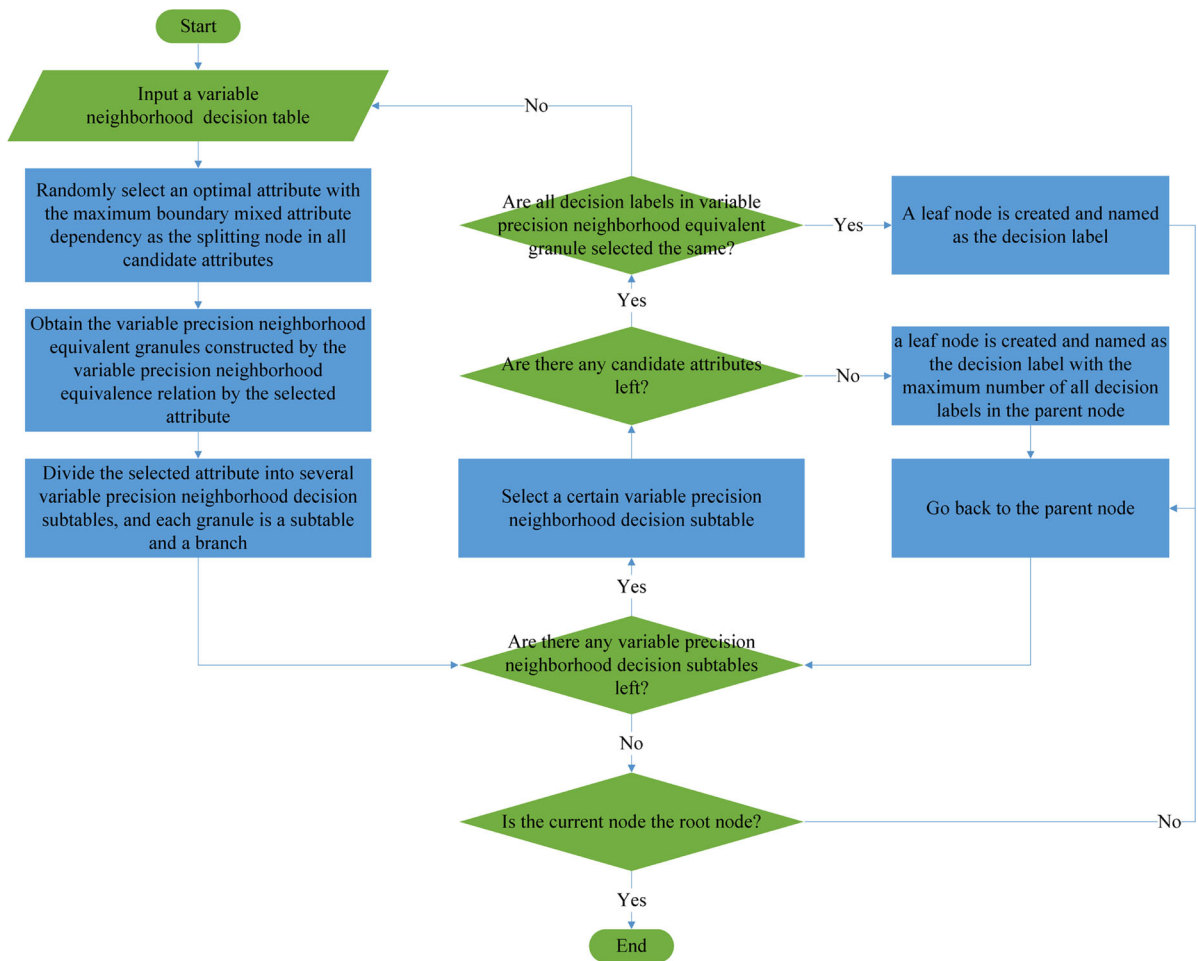


Fig. 1 The flowchart of Algorithm 1

Table 7 The information of the 14 datasets

	Dataset	No. of samples	No. of condition attributes	No. of classes
1	Crayo	90	6	2
2	Iris	150	4	3
3	wine	178	13	3
4	plrx	182	18	2
5	wpbc	194	33	2
6	Sheesegmentation	210	19	7
7	seeds	210	7	3
8	glass	214	10	6
9	heart	270	13	2
10	ecoli	336	7	7
11	ILPD	583	10	2
12	ENB2012	768	9	6
13	magic	2501	10	2
14	winequality	4899	11	7

Table 8 The running time(s)

	Dataset	AD (NRJGS)	AD (NRAGS)	AD (NROGS)	AD (NRGPS)	BMAD (NRJGS)	BMAD (NRAGS)	BMAD (NROGS)	BMAD (NRGPS)
1	Crayo	26	14	17	22	27	11	14	13
2	Iris	53	34	34	38	53	28	32	25
3	wine	296	176	151	178	253	125	160	125
4	plrx	167	114	100	102	124	81	80	86
5	wdbc	502	359	296	300	569	256	243	254
6	Sheesegmentation	1256	824	1338	1404	1267	643	783	1107
7	seeds	273	180	169	164	216	135	131	143
8	glass	527	300	316	312	427	281	250	253
9	heart	604	376	460	252	265	307	257	226
10	ecoli	2524	1368	1584	874	1431	1228	778	820
11	ILPD	1848	1264	1035	1152	1447	1118	972	824
12	ENB2012	9438	4992	5128	5852	7005	4368	4156	4172
13	magic	41136	25970	56712	24814	41360	21632	27674	21453
14	winequality	710302	384270	430668	373062	668686	365484	356276	363879
	Average	54925	30017	35572	29180	51654	28264	27986	28098

pared with $BMAD(NRJGS)$, because the neighborhood algebraic similarity of them are both Jaccard similarity.

From Table 8 and Fig. 2, it can be seen that in most datasets, the running time of Algorithm 1 is less than that of the algorithm proposed in [14]. The effectiveness of Algorithm 1 is more clear as the number of samples increases. In this case, the cost in the computation of boundary coefficient in Algorithm 1 is a worth in getting higher accuracy. In addition, we can clearly observe that the running time of Algorithm 1 is still high, which indicates that the Algorithm 1 is not appropriate for large datasets.

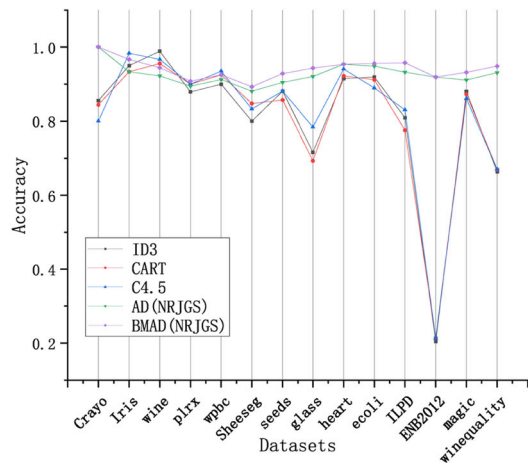
Table 9 and Fig. 3 are the accuracy comparisons. We can observe that the accuracy of Algorithm 1 is the highest, where the highest average of accuracy reaches 0.9518. Especially, the accuracy of the crayo dataset reaches 1 in all four neighborhood similarities. Compared with the three classical algorithms, the increment of accuracy reaches 10%. Compared with the algorithm in [14], the increment of accuracy reaches above 2%, even if the accuracy of all four neighborhood similarities in [14] has reached more than 93%. Hence, the improvement is significant.

Table 10 and Fig. 4 are the number of leaves comparisons between algorithms. We can see that the number of leaves of the Algorithm 1 is more than the classical algorithms, which indicates that the degree of fitting of the Algorithm 1 is higher than the classical algorithms. Compared with the algorithm in [14], it is clear that there is not much diversity between the Algorithm 1 and [14]. And it can be seen that the variation in the number of leaves is slight and acceptable. Therefore, the

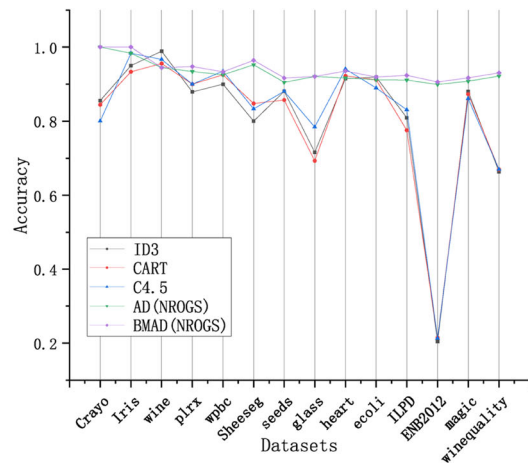
above experimental results show that Algorithm 1 is meaningful and effective.

5 Conclusion and future work

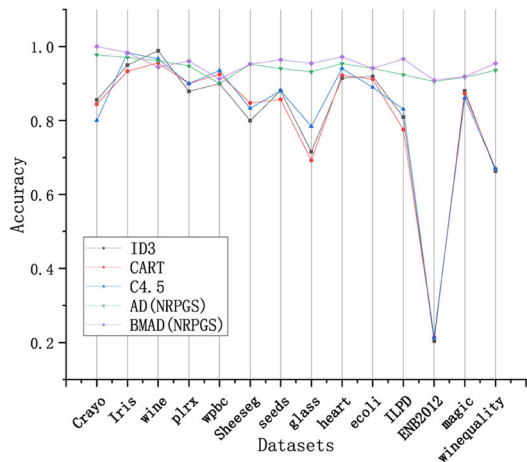
In order to consider the change of boundary region in an improved decision tree algorithm based on rough set theory, the boundary coefficient is proposed in this paper. The boundary coefficient can reflect the difficulty when we use the available knowledge to approximately represent the decision, that is, the possibility of selecting one attribute to start the partition in the decision tree algorithm. Then, the boundary mixed attribute dependency is introduced by combining the boundary coefficient with the attribute dependency as the decision rule. This decision rule has better performance in selecting attributes than the single attribute dependency in decision tree learning. The experimental results show that the accuracy of our algorithm is larger than the compared algorithms, of which the maximum reaches 0.9518. Moreover, in terms of the running time and the number of leaves, our algorithm also performs well. However, the running time is still too long. Because the Algorithm 1 needs to calculate all samples to form neighborhood granular. Then it is needed to calculate all neighborhood granular to form variable precision equivalence relation, which results in long running time and high accuracy. Therefore, further research is needed on how to reduce the running time while maintaining the high accuracy and apply the Algorithm 1 to larger datasets, such as the perspective of local.



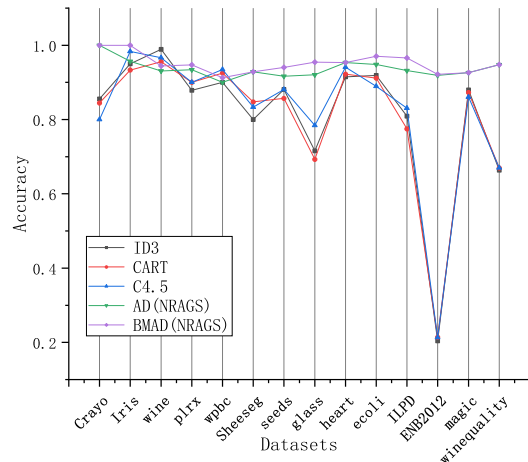
(a) AD(NRJGS) and BMAD(NRJGS)



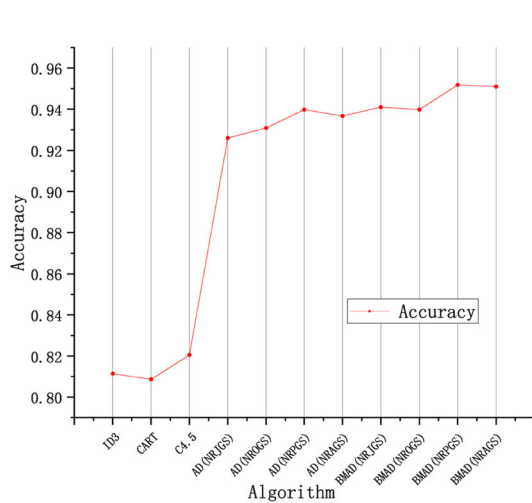
(b) AD(NROGS) and BMAD(NROGS)



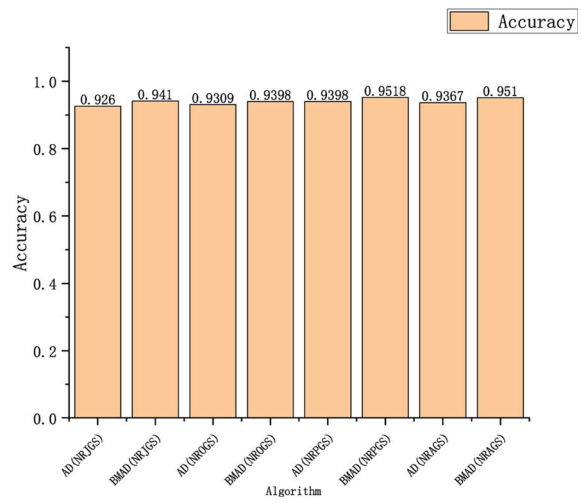
(c) AD(NRPGS) and BMAD(NRPGS)



(d) AD(NRAGS) and BMAD(NRAGS)



(e) Average

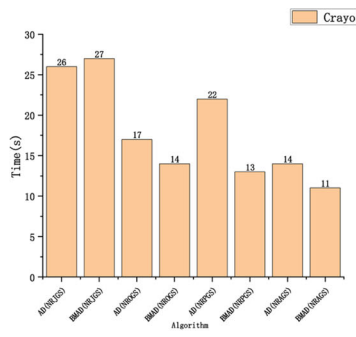


(f) Average between eight algorithms

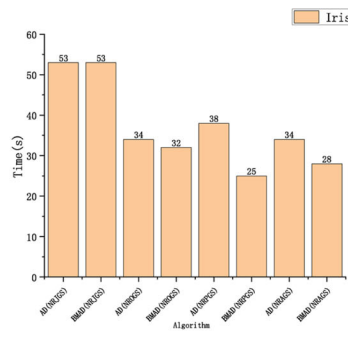
Fig. 2 The accuracy comparison between algorithms

Table 9 The accuracy

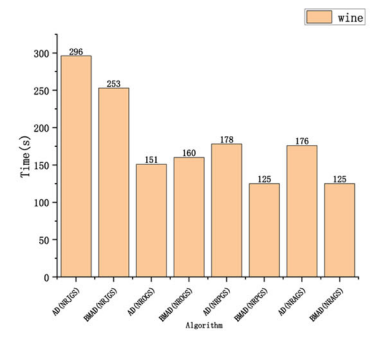
Dataset	ID3	CART	C4.5	AD(NRJGS)	AD(NRAGS)	AD(NROGS)	AD(NRGPS)	BMAD(NRJGS)	BMAD(NRAGS)	BMAD(NROGS)	BMAD(NRGPS)
1 Crayo	0.8556	0.8445	0.8	1	1	1	0.9778	1	1	1	1
2 Iris	0.95	0.9333	0.9833	0.9533	0.9567	0.9833	0.97	0.9667	1	1	0.9833
3 wine	0.9889	0.9556	0.9667	0.9222	0.9306	0.9444	0.9615	0.9444	0.9444	0.9444	0.9444
4 plrx	0.879	0.9	0.9	0.9221	0.9342	0.9342	0.9474	0.9079	0.9474	0.9474	0.9605
5 wpbc	0.9	0.925	0.935	0.9125	0.9	0.925	0.9	0.925	0.9125	0.9333	0.9125
6 Sheesgmentation	0.8	0.8476	0.8333	0.881	0.9286	0.9524	0.9524	0.8929	0.9286	0.9643	0.9524
7 seeds	0.881	0.8571	0.881	0.9048	0.9167	0.9048	0.9405	0.9286	0.9405	0.9167	0.9642
8 glass	0.7159	0.6931	0.7841	0.9205	0.9205	0.9205	0.9318	0.9432	0.9545	0.9205	0.9545
9 heart	0.9148	0.9222	0.9407	0.9537	0.9537	0.9167	0.9537	0.9537	0.9537	0.9352	0.9722
10 ecoli	0.9191	0.9117	0.8897	0.9485	0.9485	0.9118	0.9412	0.9559	0.9706	0.9191	0.9412
11 ILPD	0.8093	0.7754	0.8305	0.9322	0.9322	0.911	0.9237	0.9576	0.9661	0.9237	0.9661
12 ENB2012	0.2039	0.2143	0.213	0.9188	0.9188	0.899	0.9058	0.9188	0.922	0.9058	0.9091
13 magic	0.8801	0.873	0.8608	0.911	0.926	0.908	0.917	0.932	0.926	0.917	0.919
14 winequality	0.6638	0.6699	0.6699	0.9311	0.948	0.9217	0.9355	0.9486	0.948	0.9306	0.9546
Average	0.8115	0.8086	0.8206	0.926	0.9367	0.9309	0.9398	0.941	0.951	0.9398	0.9518



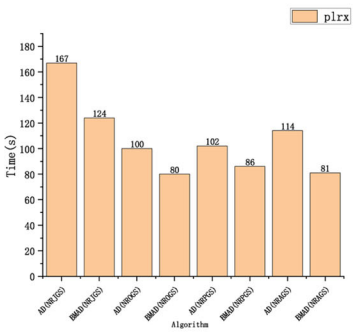
(a) Crayo



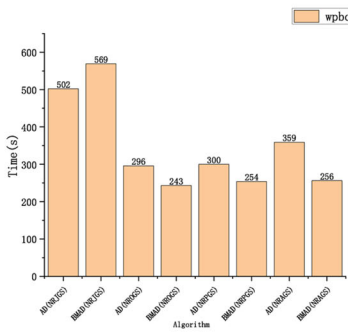
(b) Iris



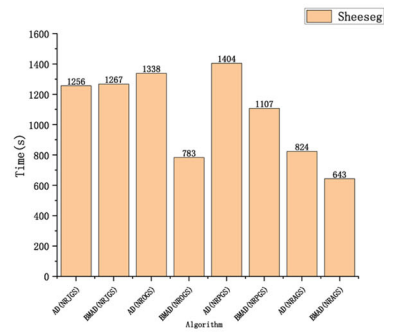
(c) wine



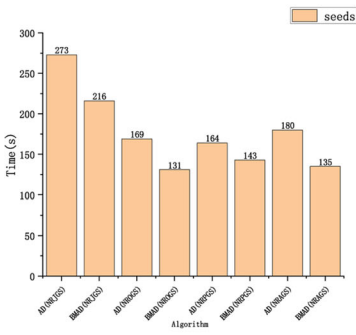
(d) plrx



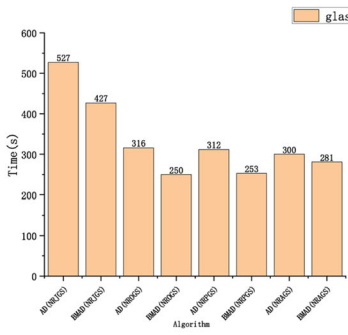
(e) wpbc



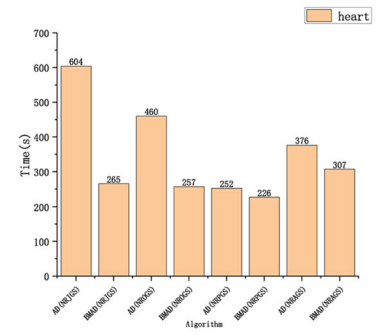
(f) Sheesegmentation



(g) seeds

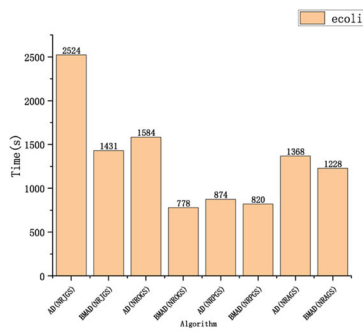


(h) glass

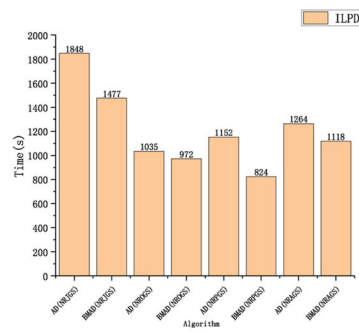


(i) heart

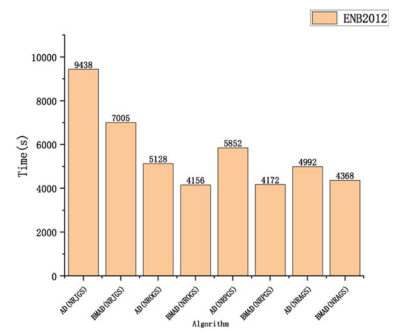
Fig. 3 The running time comparison between algorithms



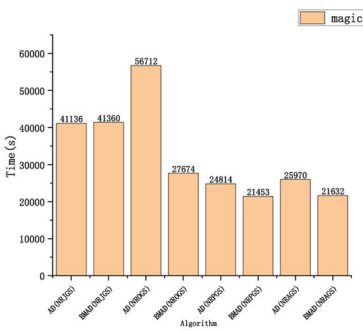
(j) ecoli



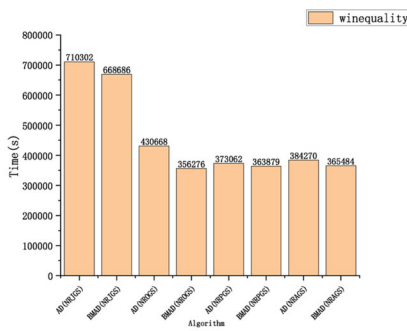
(k) ILPD



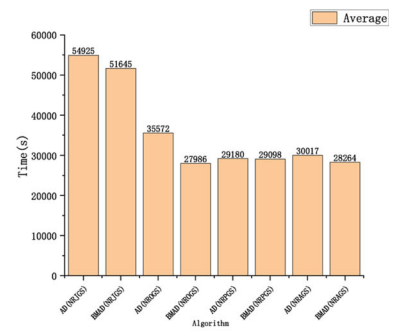
(l) ENB2012



(m) magic



(n) winequality

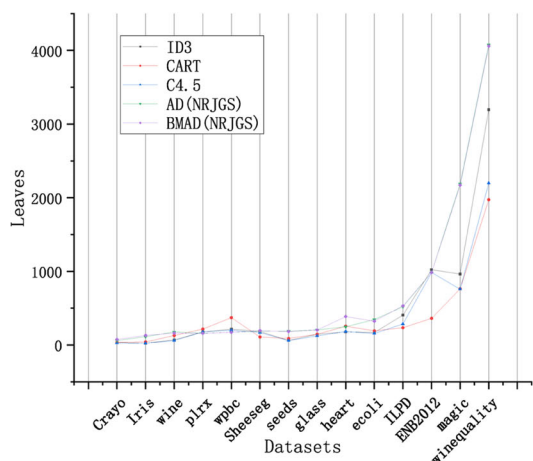


(o) Average

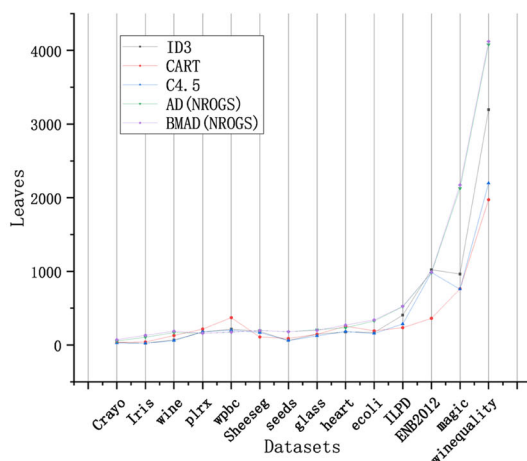
Fig. 3 continued

Table 10 The number of leaves

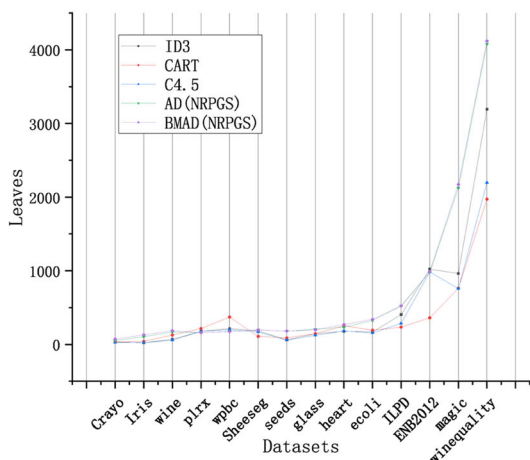
Dataset	ID3	CART	C4.5	AD(NRJGS)	AD(NRAGS)	AD(NROGS)	AD(NRGPS)	BMAD(NRJGS)	BMAD(NRAGS)	BMAD(NROGS)	BMAD(NRGPS)
1 Crayo	32	35	27	60	53	75	62	74	58	71	62
2 Iris	23	40	24	113	106	110	104	130	102	131	98
3 wine	69	128	59	176	168	171	158	161	166	186	164
4 plrx	178	216	178	159	159	159	159	158	161	160	159
5 wpbc	215	372	200	172	173	169	170	172	172	175	175
6 Sheesegmentation	184	108	169	188	197	190	195	193	200	196	192
7 seeds	61	89	59	186	179	184	178	182	181	180	182
8 glass	145	145	124	205	207	204	202	203	205	197	204
9 heart	178	255	182	247	233	253	251	385	259	270	257
10 ecoli	169	192	157	344	326	336	326	324	338	341	325
11 ILPD	405	234	285	519	520	517	537	531	510	525	537
12 ENB2012	1023	361	986	984	980	961	947	984	897	989	985
13 magic	963	764	758	2190	2126	2115	2213	2170	2174	2173	2071
14 winequality	3194	1973	2196	4076	4080	4079	4112	4058	4065	4121	4068
Average	488	350	385	686	679	679	686	694	693	677	677



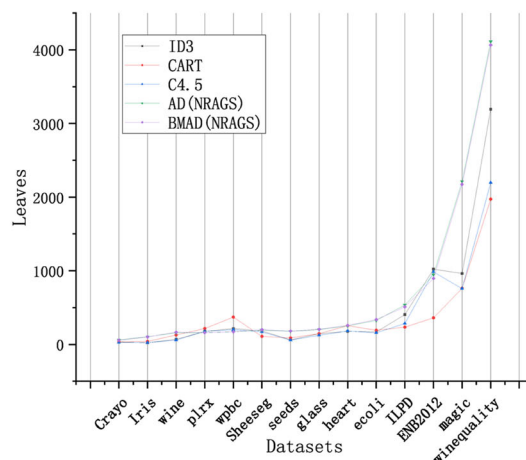
(a) AD(NRJGS) and BMAD(NRJGS)



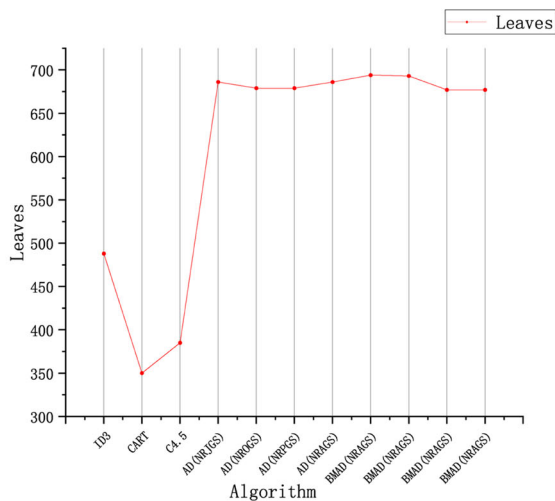
(b) AD(NROGS) and BMAD(NROGS)



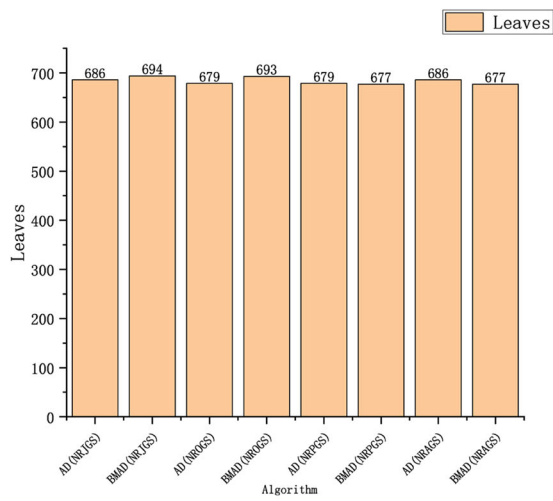
(c) AD(NRPGS) and BMAD(NRPGS)



(d) AD(NRAGS) and BMAD(NRAGS)



(e) Average



(f) Average between eight algorithms

Fig. 4 The number of leaves comparison between algorithms

Acknowledgements This research is supported by the National Natural Science Foundation of China under Grant Nos. 62166001, Graduate Innovation Funding Program of Gannan Normal University, China under Grant No. YCX22A025.

Author Contributions **Bowen Lin:** Conceptualization, Methodology, Writing - original draft, Software, Validation, Data-curation. **Caihui Liu:** Conceptualization, Methodology, Writing - review & editing, Validation, Supervision. **Duoqian Miao:** Writing - review & editing.

Declarations

Competing Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Breiman L (2017) Classification and regression trees. Routledge
- Broelemann K, Kasneci G (2019) A gradient-based split criterion for highly accurate and transparent model trees. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence. International Joint Conferences on Artificial Intelligence Organization, IJCAI-2019, pp 2030–2037. <https://doi.org/10.24963/ijcai.2019/281>
- Gao C, Zhou J, Miao D et al (2021) Granular-conditional-entropy-based attribute reduction for partially labeled data with proxy labels. *Inf Sci* 580:111–128. <https://doi.org/10.1016/j.ins.2021.08.067>
- Gao C, Li Y, Zhou J et al (2022) Global structure-guided neighborhood preserving embedding for dimensionality reduction. *Int J Mach Learn Cybern* 13(7):2013–2032. <https://doi.org/10.1007/s13042-021-01502-6>
- Gao C, Wang Z, Zhou J (2022) Three-way approximate reduct based on information-theoretic measure. *Int J Approx Reason* 142:324–337. <https://doi.org/10.1016/j.ijar.2021.12.008>
- Han X, Zhu X, Pedrycz W et al (2023) A three-way classification with fuzzy decision trees. *Appl Soft Comput* 132:109788. <https://doi.org/10.1016/j.asoc.2022.109788>
- Hu Q, Yu D, Liu J et al (2008) Neighborhood rough set based heterogeneous feature subset selection. *Inf Sci* 178(18):3577–3594. <https://doi.org/10.1016/j.ins.2008.05.024>
- Hu Q, Yu D, Xie Z (2008) Neighborhood classifiers. *Expert Syst Appl* 34(2):866–876. <https://doi.org/10.1016/j.eswa.2006.10.043>
- Jiang F, Sui Y, Cao C (2013) An incremental decision tree algorithm based on rough sets and its application in intrusion detection. *Artif Intell Rev* 40(4):517–530. <https://doi.org/10.1007/s10462-011-9293-z>
- Kang Y, Dai J (2023) Attribute reduction in inconsistent grey decision systems based on variable precision grey multigranulation rough set model. *Appl Soft Comput* 133:109928. <https://doi.org/10.1016/j.asoc.2022.109928>
- Laber E, Murtinho L, Oliveira F (2023) Shallow decision trees for explainable k-means clustering. *Pattern Recogn* 137:109239. <https://doi.org/10.1016/j.patcog.2022.109239>
- Lin T (1997) Neighborhood systems—a qualitative theory for fuzzy and rough sets. *Adv Mach Intell Soft Comput* 4:132–155
- Lin TY (2003) Neighborhood systems: mathematical models of information granulations. In: SMC’03 Conference proceedings. 2003 IEEE international conference on systems, man and cybernetics. conference theme - system security and assurance (Cat. No.03CH37483), pp 3188–3193 vol.4. <https://doi.org/10.1109/ICSMC.2003.1244381>
- Liu C, Lin B, Lai J et al (2022) An improved decision tree algorithm based on variable precision neighborhood similarity. *Inf Sci* 615:152–166. <https://doi.org/10.1016/j.ins.2022.10.043>
- Liu C, Lai J, Lin B et al (2023) An improved id3 algorithm based on variable precision neighborhood rough sets. *Appl Intell* 53:23641–23654. <https://doi.org/10.1007/s10489-023-04779-y>
- Luo C, Cao Q, Li T et al (2023) Mapreduce accelerated attribute reduction based on neighborhood entropy with apache spark. *Expert Syst Appl* 211:118554. <https://doi.org/10.1016/j.eswa.2022.118554>
- Luo C, Wang S, Li T et al (2023) Rhdfos: a distributed online algorithm towards scalable streaming feature selection. *IEEE Trans Parallel and Distrib Syst* 34(6):1830–1847. <https://doi.org/10.1109/TPDS.2023.3265974>
- Ma Z, Mi J (2016) Boundary region-based rough sets and uncertainty measures in the approximation space. *Inf Sci* 370–371:239–255. <https://doi.org/10.1016/j.ins.2016.07.040>
- Mani A (2018) Algebraic methods for granular rough sets. Springer International Publishing, Cham, pp 157–335. https://doi.org/10.1007/978-3-030-01162-8_3
- Mani A (2022) Granularity and rational approximation: rethinking graded rough sets. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 33–59. https://doi.org/10.1007/978-3-662-66544-2_4
- Miao D, Wang J (1997) Rough sets based approach for multivariate decision tree construction. *Chin J Softw* 8(6):26–32(In Chinese with English Abstract)
- Parthaláin N, Shen Q, Jensen R (2010) A distance measure approach to exploring the rough set boundary region for attribute reduction. *IEEE Trans Knowl Data Eng* 22(3):305–317. <https://doi.org/10.1109/TKDE.2009.119>
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11(5):341–356. <https://doi.org/10.1007/BF01001956>
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1:81–106
- Quinlan JR (2014) C4. 5: programs for machine learning. Elsevier
- Ren Y, Zhu X, Bai K et al (2023) A new random forest ensemble of intuitionistic fuzzy decision trees. *IEEE Trans Fuzzy Syst* 31(5):1729–1741. <https://doi.org/10.1109/TFUZZ.2022.3215725>
- Wang J, Qian Y, Li F et al (2020) Fusing fuzzy monotonic decision trees. *IEEE Trans Fuzzy Syst* 28(5):887–900. <https://doi.org/10.1109/TFUZZ.2019.2953024>
- Xie X, Xianyong Z, Wuanye W et al (2021) Neighborhood decision tree construction algorithm based on variable precision neighborhood equivalent granules. *Chin J Comput Appl* 42(2):382–388 (In Chinese with English Abstract)
- Xu B, Zhang X, Feng S (2018) Weighted dependence of neighborhood rough sets and its heuristic reduction algorithm. *Chin Pattern Recognit Artif Intell* 31(3):256–264 (In Chinese with English Abstract)
- Xu W, Yuan Z, Liu Z (2023) Feature selection for unbalanced distribution hybrid data based on k-nearest neighborhood rough set. *IEEE Trans Artif Intell* pp 1–15. <https://doi.org/10.1109/TAI.2023.3237203>
- Yang X, Chen Y, Fujita H et al (2022) Mixed data-driven sequential three-way decision via subjective-objective dynamic fusion. *Knowl Based Syst* 237:107728. <https://doi.org/10.1016/j.knosys.2021.107728>
- Yang X, Li M, Fujita H et al (2022) Incremental rough reduction with stable attribute group. *Inf Sci* 589:283–299. <https://doi.org/10.1016/j.ins.2021.12.119>
- Yao Y, Zhang X, Chen S et al (2021) Decision-tree induction algorithm based on attribute purity degree. *Chinese Comput Eng Des* 42(1):142–149 (In Chinese with English Abstract)

34. Zhai J, Wang X, Zhang S et al (2018) Tolerance rough fuzzy decision tree. *Inf Sci* 465:425–438. <https://doi.org/10.1016/j.ins.2018.07.006>
35. Zhang X, Yao Y (2022) Tri-level attribute reduction in rough set theory. *Expert Syst Appl* 190:116187. <https://doi.org/10.1016/j.eswa.2021.116187>
36. Zhang X, Yuan Z, Miao D (2023) Outlier detection using three-way neighborhood characteristic regions and corresponding fusion measurement. *IEEE Trans Knowl Data Eng* 1–14. <https://doi.org/10.1109/TKDE.2023.3312108>
37. Ziarko W (1993) Variable precision rough set model. *J Comput Syst Sci* 46(1):39–59. [https://doi.org/10.1016/0022-0000\(93\)90048-2](https://doi.org/10.1016/0022-0000(93)90048-2)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Bowen Lin received his bachelor's degree from Gannan Normal University in 2021. He is currently pursuing the master's degree with the School of Mathematics and Computer Science, Gannan Normal University. His research interests include granular computing and machine learning.



Caihui Liu received his Ph.D. degree from Tongji University, China, in 2013. He is currently a Professor with the Department of Mathematics and Computer Science, Gannan Normal University, China. His research interests include artificial intelligence, machine learning, and granular computing.



Duoqian Miao is currently a Professor and a Ph.D. Tutor with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. He has published more than 200 papers in *IEEE TRANSACTIONS ON CYBERNETICS*, *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, *Pattern*

Recognition, *Information Sciences*, and *Knowledge-Based Systems*. His interests include machine learning, data mining, big data analysis, granular computing, artificial intelligence, and text image processing. Dr. Miao's representative awards include the Second Prize of Wuwenjun AI Science and Technology in 2018, the First Prize of Natural Science of Chongqing in 2010, the First Prize of Technical Invention of Shanghai in 2009, and the First Prize of the Ministry of Education Science and Technology Progress Award in 2007. He serves as the Honorary Chair of the CAAI Granular Computing Knowledge Discovery Technical Committee, the Vice-Director for MOE Key Laboratory of Embedded System & Service Computing, and the Vice-President of the Shanghai Association for Artificial Intelligence. He serves as an Associate Editor for *International Journal of Approximate Reasoning* and *Information Sciences* and an Editor of the *Journal of Computer Research and Development* (in Chinese).