Information Sciences 615 (2022) 152–166

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

An improved decision tree algorithm based on variable precision neighborhood similarity

Caihui Liu [a,*], Bowen Lin [a], Jianying Lai [a], Duoqian Miao [b]

[a] School of Mathematics and Computer Science, Gannan Normal University, Ganzhou 341000, Jiangxi, China
[b] Department of Computer Science and Technology, Tongji University, 201804 Shanghai, China

ARTICLE INFO

Article history:
Received 20 July 2022
Received in revised form 30 August 2022
Accepted 6 October 2022
Available online 10 October 2022

Keywords:
Decision tree algorithm
Neighborhood geometry similarity
Neighborhood algebraic similarity
Equivalence relation
Variable precision neighborhood rough set
Attribute dependence

ABSTRACT

The decision tree algorithm has been widely used in data mining and machine learning due to its high accuracy, low computational cost and high interpretability. However, when dealing with the continuous data, the classical decision tree algorithm needs to replace continuous attributes with discretized attributes by the strategy of discretization. Discretization may cause a loss of information structure, which will affect the performance of classification. To tackle this problem, many researchers have proposed different decision tree methods based on variable precision neighborhood rough sets. However, these methods do not consider the geometric structure of neighborhood systems, which may lead to a contradiction in the transitivity of the equivalence relation. In this paper, we first define a novel neighborhood geometric similarity in a neighborhood system from the perspective of geometry. Second, by combining the neighborhood geometric similarity and the neighborhood algebraic similarity, we propose four new kinds of neighborhood similarities, which can solve the contradictory transitivity of the equivalence relation. Third, a variable precision neighborhood rough set model is constructed using the new similarities, and a novel decision tree algorithm is proposed based on this model, where the degree of attribute dependence is used as the partition measure. Experimental results on 14 selected datasets from the UCI Machine Learning Repository show that our algorithm is effective. The average accuracy of our algorithm is over 90%, which is 10% higher than the classical decision tree algorithms, and the number of leaf nodes increases slightly.

1. Introduction

Classification has always been a popular field of study in data mining and machine learning. In classification algorithms, the decision tree algorithm is extensively applied to many areas for its high accuracy, low computational cost and high interpretability. The process of building a decision tree can be summarised as the following three steps: (1) select the best attribute (called 'splitting attribute') according to a certain partition measure; (2) divide the training set with the selected splitting attribute; (3) produce a branch corresponding to each classification of the splitting attribute. The algorithm is recursively applied to each classification derived from the splitting attribute. If all samples in a certain classification come from the same category, then a leaf node with the name of the category is generated [1]. From the construction of the decision tree algorithm, it is clearly that each step of the decision tree algorithm depends on the partition measure. Therefore, for the deci-

* Corresponding author.
E-mail addresses: liucaihui@gnnu.edu.cn (C. Liu), dqmiao@tongji.edu.cn (D. Miao).

https://doi.org/10.1016/j.ins.2022.10.043
0020-0255/

sion tree algorithm, the choose of the partition measure directly determines its the effect. The representative decision tree algorithms include: the CART algorithm [2] on the basis of the Gini index, the ID3 algorithm [3] on the basis of information gain, and the C4.5 algorithm [4] on the basis of information gain ratio, where C4.5 is an improvement of ID3. Although the decision tree algorithms have achieved great success, they are unable to handle the continuous data directly, which dramatically limits the field of practical applications. When regarding the continuous data, these algorithms usually use the strategy of discretization [5]. However, discretization may result in a loss of information structure. Since the discretization of continuous attribute does not consider the membership degree between continuous values and discrete values, the discretization results may lose at least two types of structures: structure of neighborhood and ordered structure in reality space [6]. For instance, if we are aware of the distance among two samples, then we can obtain the degree of closeness of samples in real space. However, this type of information will disappear after the discretization. Hence, the data discretization may affect the performance of a decision tree algorithm.

Rough set theory introduced by Pawlak [7] is a method extensively used in attribute reduction [8–11] and classification [12,13]. In rough set theory, the set of research targets is denoted as an universe. An universe is divided into several exclusive equivalent classes through an equivalence relation [5]. These equivalent classes can be used to describe any concept in the universe. But the standard definition of the set inclusion relation in Pawlak rough set model is too rigorous to contain some degree of error classification. The Pawlak rough set model is an excellent model for dealing with the discrete data, but when it deals with the continuous data, a discretization process should be employed. In [14,15], Hu et al. proposed neighbourhood rough sets by the introduction of neighbourhood system. This model can directly deal with the continuous data without the process of discretization, which can effectively solve the problem of Pawlak rough set model for dealing with the continuous data.

With the deepening of the progress of research on the rough sets [16–18], there is a growing body of experts and scholars that recognises the consistence of the rough set theory and the decision tree model. They combined rough set theory with decision tree algorithms and obtained good results for different application problems. For instance, Jiang et al. [19] presented an incremental decision tree algorithm on the basis of rough set theory. Moshkov [20] investigated decision trees on the basis of the results of test theory and rough set theory.

The degree of attribute dependency in rough set theory is a description of the dependency of one attribute on another attribute, which is the same idea as the classical decision tree algorithm for selecting the splitting attributes. Therefore, many scholars have applied the degree of attribute dependency and various improved measures of it as the partition measures in decision tree algorithms and obtained good results. For example, Wang et al. [21] proposed a decision tree algorithm on the basis of attribute importance and variable precision rough sets [22]. Ye [23] proposed a new knowledge-dependent decision tree algorithm by improving attribute dependency. However, all of the above improvements on the basis of the Pawlak rough set model can only handle the discrete data. An efficient approach to handle continuous data with the Pawlak rough set model is to discretize the data, but this approach may result in a loss of information structure. To avoid the loss of information structure, many researchers have used the neighborhood rough set model, which can directly handle continuous data without discretization. Once the Pawlak rough set model is replaced by the neighbourhood rough set model that can directly handle continuous data, the definition of equivalence relation is a hurdle that cannot be bypassed if we use the attribute dependence as the partition measure. Although the definitions of neighborhood rough set were proposed for more than a decade, the definitions of equivalence relations for neighborhood rough set model have not received much attention [15]. Xie and Zhang [24] pointed out that the neighborhood similarity between two neighborhood granules can be defined by introducing the Jaccard similarity into the neighborhood system. Based on this observation, they defined the variable precision neighborhood equivalence classes by the neighborhood similarity and the variable precision threshold β. Furthermore, they replaced the attribute classes in the CART algorithm by the variable precision neighborhood equivalence classes of each attribute to calculate the gini index, which can improve the accuracy of the algorithm greatly. However, the neighborhood similarity proposed in [24] only focuses on the algebraic structure of the neighborhood system, that is, the magnitude relationship of the similarity values between two neighborhood granules. In fact, in a neighborhood system the neighborhood granules not only have algebraic structure but also geometric structure. We found that if the definition of the neighborhood similarity between two neighborhood granules does not take account into the geometric structure, the transitivity of the variable precision equivalence relation would be contradictory.

In order to solve the problems above, we firstly define a neighborhood geometric similarity among two neighborhood granules and use the neighborhood geometric similarity to constrain the neighborhood algebraic similarity [25], then proposing a new neighborhood similarity and extending three new similarity measures by introducing the pessimistic, optimistic and average ideas [26–28] in the Jaccard similarity. Second, the variable precision neighborhood equivalence relations are defined based on the neighborhood similarity by introducing variable precision thresholds [5,29–31]. According to the above work, the contradictory transitivity of the equivalence relation can be solved. Third, the variable precision neighborhood rough set model is constructed and a novel decision tree algorithm is proposed based on this model, where the attribute dependency is used as the partition measure. The experimental performance show that the average improvement of the accuracy of the algorithm in this paper is greater than 10% and the number of leaf nodes increases slightly.

The main contributions of this paper are given as follows. (1) We define the neighborhood geometric similarity by considering the geometric structure relationship between two neighborhood granules and combine it with the other four neighborhood algebraic similarities to construct the neighborhood similarity. (2) We define a new variable precision neighborhood rough set model by means of the new neighborhood similarity and variable precision threshold. (3) By using

the attribute dependence in the variable precision rough set model as the partition measure, we construct a decision tree algorithm, which provides an effective strategy to handle continuous data.

The organization of the remaining part of this paper is as follows. Section 2 mainly reviews the basic knowledge of variable precision rough set model, neighborhood rough sets model, similarity measure and decision tree algorithm. Section 3 introduces how to combine similarity measure with neighborhood rough sets, and the philosophy of variable precision is introduced to improve the decision tree algorithm. Section 4 proves the feasibility of the proposed algorithm by simulation analysis. Finally, we conclude our work.

2. Preliminaries

2.1. Neighborhood rough set model

Neighborhood rough set model is introduced by Hu et al. [15], which is dedicated to tackle the problem of data loss caused by data discretization when processing continuous data using Pawlak rough set model.

For any two samples $x^i = (x_1^i, x_2^i, \ldots, x_H^i) \in U, x^j = \left(x_1^j, x_2^j, \ldots, x_H^j\right) \in U, B$is a feature space and $B \subseteq C$, the Minkowsky distance between x^i and x^j under B is defined as follows.

$$DF^B(x^i, x^j) = \left(\sum_{h=1}^{H} \left|x_h^i - x_h^j\right|^p\right)^{\frac{1}{p}} \tag{1}$$

where DF is a distance measure. For $\forall x, y, z \in U, DF$ usually satisfies the following conditions [15]:

(1) $DF(x, y) = 0 \quad \text{iff} \quad x = y$;
(2) $DF(x, y) = DF(y, x)$;
(3) $DF(x, z) \leqslant DF(x, y) + DF(y, z)$. Convenient for calculation, in this paper we keep $p = 1$ (also called the Manhattan distance).

Let $\forall\, x^i \in U, B \subseteq C$, thus the neighborhood of x^i in the attribute set B can be defined as follows [15].

$$\delta^B(x^i) = \left\{x^j | DF^B(x^i, x^j) \leqslant \delta, \right\}. \tag{2}$$

Let U, N be a neighborhood approximation space, where U is the universe and N is the neighborhood relation.

1. For $X \subseteq U$, the lower approximation and the upper approximation of the set X can be defined as follows [15].

$$\underline{N}X = \left\{x^i \in U | \delta(x^i) \subseteq X\right\}. \tag{3}$$

$$\overline{N}X = \left\{x^i \in U | \delta(x^i) \cap X \neq \emptyset\right\}. \tag{4}$$

2. For $X \subseteq U$, the boundary region of the set X is given as follows [15].

$$BNR(X) = \overline{N}X - \underline{N}X. \tag{5}$$

2.2. Similarity

The similarity measure is considered as one of the popular tools to describe the degree of similarity among two objects. By comparing the similarity among objects, we can get the classification results of objects or reveal the hidden information of objects [26]. Among the various similarity measures, Jaccard similarity is extensively used to assess the similarity between two given sets [27]. In this paper, we concentrate on Jaccard similarity and its variants.

1. The Jaccard similarity between set X_1 and set X_2 is defined as follows [26].

$$J_{(X_1, X_2)} = \frac{card(X_1 \cap X_2)}{card(X_1 \cup X_2)}. \tag{6}$$

2. When we bring the optimistic influence into the Jaccard similarity, the Optimistic similarity between set X_1 and set X_2 can be defined as follows [25].

$$O_{(X_1, X_2)} = \frac{card(X_1 \cap X_2)}{\min\left(card(X_1), card(X_2)\right)}. \tag{7}$$

3. When we bring the pessimistic influence into the Jaccard similarity, the Pessimistic similarity between set X_1 and set X_2 can be defined as follows [25].

$$P_{(X_1X_2)} = \frac{card(X_1 \cap X_2)}{\max(card(X_1), card(X_2))}. \tag{8}$$

4. When we bring the average thought into the Jaccard similarity, the Average similarity between set X_1 and set X_2 can be defined as follows.

$$A_{(X_1X_2)} = \frac{card(X_1 \cap X_2)}{average(card(X_1), card(X_2))}. \tag{9}$$

3. An improved decision tree algorithm based on variable precision neighborhood similarity

In order to solve the problem that the decision tree algorithm needs to discretize data when dealing with continuous data, we use the variable precision neighborhood rough set model to construct a decision tree algorithm, which can deal with continuous data without discretization. In this section, we define the neighborhood geometric similarity and four kinds of neighborhood algebraic similarities, and combine them to form four new kinds of neighborhood similarities. The new neighborhood similarities can not only measure the similarity between two neighborhood granulations, but also avoid the transitivity contradiction of the equivalence relation defined later. The variable precision neighborhood equivalence relation is jointly defined by neighborhood similarity and variable-precision threshold. From this, a variable precision neighborhood rough set model is constructed. Finally, the decision tree algorithm is constructed by using the degree of attribute dependency in the variable precision neighborhood rough set model as the partition measure.

3.1. Neighborhood similarity

The partition of continuous attributes by using rigorous equivalence relation would lead to problems such as too many equivalence classes and poor partition effect. In this subsection we first introduce four similarity measures into the neighborhood rough set model to define neighborhood algebraic similarities. Then we use neighborhood algebraic similarity and neighborhood geometric similarity to define new neighborhood similarities.

Definition 1. Let $NDS = \langle U, C \cup D\rangle$ be a neighborhood decision system, B is a attribute set, $B \subseteq C$ and $x^1, x^2 \in U$. Four neighborhood algebraic similarities between neighborhood granule $\delta^B(x)$ and neighborhood granule $\delta^B(y)$ can be defined as follows.

1. The neighborhood Jaccard similarity is defined as follows [24].

$$J_{(\delta^B(x^1),\delta^B(x^2))} = \frac{|\delta^B(x^1) \cap \delta^B(x^2)|}{|\delta^B(x^1) \cup \delta^B(x^2)|}. \tag{10}$$

2. The neighborhood Optimistic similarity is defined as follows.

$$O_{(\delta^B(x^1),\delta^B(x^2))} = \frac{|\delta^B(x^1) \cap \delta^B(x^2)|}{\min(|\delta^B(x^1)|, |\delta^B(x^2)|)}. \tag{11}$$

3. The neighborhood Pessimistic similarity is defined as follows.

$$P_{(\delta^B(x^1),\delta^B(x^2))} = \frac{|\delta^B(x^1) \cap \delta^B(x^2)|}{\max(|\delta^B(x^1)|, |\delta^B(x^2)|)}. \tag{12}$$

4. The neighborhood Average similarity is defined as follows.

$$A_{(\delta^B(x^1),\delta^B(x^2))} = \frac{|\delta^B(x^1) \cap \delta^B(x^2)|}{average(|\delta^B(x^1)|, |\delta^B(x^2)|)}. \tag{13}$$

Property 1. *Let $NDS = \langle U, C \cup D\rangle$ be a neighborhood decision system, Bis an attribute set, $B \subseteq C$and $x^1, x^2 \in U$, we have*

(1) $J_{(\delta^B(x^1),\delta^B(x^2))} \in [0,1], O_{(\delta^B(x^1),\delta^B(x^2))} \in [0,1], P_{(\delta^B(x^1),\delta^B(x^2))} \in [0,1], A_{(\delta^B(x^1),\delta^B(x^2))} \in [0,1]$;
(2) $J_{(\delta^B(x^1),\delta^B(x^2))} \leqslant P_{(\delta^B(x^1),\delta^B(x^2))} \leqslant A_{(\delta^B(x^1),\delta^B(x^2))} \leqslant O_{(\delta^B(x^1),\delta^B(x^2))}$;

(3) $\delta^B(x^1) = \delta^B(x^2) \Longleftrightarrow J_{(\delta^B(x^1),\delta^B(x^2))} = P_{(\delta^B(x^1),\delta^B(x^2))} = A_{(\delta^B(x^1),\delta^B(x^2))} = O_{(\delta^B(x^1),\delta^B(x^2))} = 1.$

Proof.

(1) Obviously, for these four formulas, when two neighboring particles do not intersect, the similarity value is 0; when two neighboring particles intersect, the similarity value is greater than 0. At this time, the numerator takes the number of elements of the small neighboring particles as the maximum value, and the denominator takes the number of elements of the small neighboring particles as the minimum value, at this time, the numerator and denominator are equal, so the maximum value of similarity is 1.
(2) For these four formulas, the numerators are the same and the denominators are arranged in size as $\min(\delta^B(x^1), \delta^B(x^2)) < average(\delta^B(x^1), \delta^B(x^2)) < \max(\delta^B(x^1), \delta^B(x^2)) < \delta^B(x^1) \cup \delta^B(x^2)$.
(3) When $\delta^B(x^1) = \delta^B(x^2)$, obviously, for all four formulas, the numerator and denominator are the same, so the four formulas take equal values.

Property 1 mainly reflects whether the four similarity measures conform to the basic properties of similarity measures and the relationship between them. From Property 1, we can know that the four similarity measures take values in the range of [0,1], and their similarity is 1 when the two neighborhood granules are equal, so they meet the definition of similarity. And these four similarity measures have a size relationship, so that it is also convenient to observe the impact of different sizes of similarity measures on the accuracy rate when comparing the accuracy rate. □

Example 1. A dataset is given as Table 1, where U is the sample set, and c^1 is a continuous attribute. Suppose that $\delta = 0.5$ and $\beta = 0.8$, then the neighborhood granules induced by c^1 are shown as Table 2. According to the neighborhood Jaccard similarity, the similarities between different neighborhood granules are shown as Table 3.

In order to relax the rigorous equivalence relation, if the neighborhood Jaccard similarity between two neighborhood granules is greater than the given precision threshold $\beta = 0.8$, then we think they are equivalent. From Table 3, we can see that the neighborhood Jaccard similarity between $\delta^{c^1}(ax^1)$ and $\delta^{c^1}(ax^4)$ is 0.8571, which is greater than 0.8, hence they are equivalent. Similarly, $\delta^{c^1}(ax^1)$ and $\delta^{c^1}(ax^6))$ are equivalent. According to the transitivity of equivalence relations, we know that $\delta^{c^1}(ax^4)$ and $\delta^{c^1}(ax^6)$ should also be equivalent. However, the neighborhood Jaccard similarity between $\delta^{c^1}(ax^4)$ and $\delta^{c^1}(ax^6)$ is only 0.7142, which is lower than 0.8. This is contradictory. From Table 4, we know that sample ax^5 is equivalent to sample ax^6 and sample ax^4 is equivalent to sample ax^7. The two classes of equivalence are not equivalent. But in the classes of equivalence of sample ax^1, they are equivalent. Hence the definition of the variable precision neighborhood equivalence by the neighborhood algebraic similarity is problematic.

Example 1 shows that if we only introduce the neighborhood algebraic similarity into the definition of neighborhood equivalence, sometimes this introduction has problems. Specifically, the definition does not take into account the geometric structure, that is, the distance between two samples.

Let us use four figures in Example 2 to illustrate this problem specifically.

Example 2. Suppose that $\delta = 5.5$ and the variable precision threshold $\beta = 0.8$. The 11 samples $A, B, C, D, E, F, G, H, I, J, K$ are evenly distributed on the line LM as shown in Figure a. The neighborhood granulation results of samples F, G and H are shown in Figure b. In this case, $\delta(F) = \{A, B, C, D, E, F, G, H, I, J, K\}, \delta(G) = \{B, C, D, E, F, G, H, I, J, K\}, \delta(H) = \{C, D, E, F, G, H, I, J, K\}$. Then let us compute the neighborhood Jaccard similarity between them. $J_{(\delta(F),\delta(G))} = 0.9090, J_{(\delta(G),\delta(H))} = 0.9$. According to the transitivity of the equivalence relation, $\delta(F)$ and $\delta(H)$ should be equivalent, too. In fact, they are equivalent since $J_{(\delta(F),\delta(H))} = 0.8181$. In this case, there is no problem in the transitivity of the equivalence relation. Because the distance between samples F and H is not large enough to cause influence.

Table 1
A given dataset.

U	c^1
ax^1	0.4
ax^2	0.9
ax^3	1
ax^4	0.5
ax^5	0.4
ax^6	0.3
ax^7	0.6

Table 2
The results of neighborhood granulation.

U	$\delta^{c^1}(x)$
ax^1	$\{ax^1, ax^2, ax^4, ax^5, ax^6, ax^7\}$
ax^2	$\{ax^1, ax^2, ax^3, ax^4, ax^5, ax^7\}$
ax^3	$\{ax^2, ax^3, ax^4, ax^7\}$
ax^4	$\{ax^1, ax^2, ax^3, ax^4, ax^5, ax^6, ax^7\}$
ax^5	$\{ax^1, ax^2, ax^4, ax^5, ax^6, ax^7\}$
ax^6	$\{ax^1, ax^4, ax^5, ax^6, ax^7\}$
ax^7	$\{ax^1, ax^2, ax^3, ax^4, ax^5, ax^6, ax^7\}$

Table 3
The calculation results of neighborhood Jaccard similarities.

Element	Neighborhood Jaccard similarity	Element	Neighborhood Jaccard similarity
$(1,1)$	1	$(2,7)$	0.8571
$(1,2)$	0.7142	$(3,4)$	0.5714
$(1,3)$	0.4285	$(3,5)$	0.4285
$(1,4)$	0.8571	$(3,6)$	0.2857
$(1,5)$	1	$(3,7)$	0.5714
$(1,6)$	0.8333	$(4,5)$	0.8571
$(1,7)$	0.8571	$(4,6)$	0.7142
$(2,3)$	0.6666	$(4,7)$	1
$(2,4)$	0.8571	$(5,6)$	0.8333
$(2,5)$	0.7142	$(5,7)$	0.8571
$(2,6)$	0.5714	$(6,7)$	0.7142

Table 4
The calculation results of equivalence classes.

Samples	Equivalence classes
ax^1	ax^4, ax^5, ax^6, ax^7
ax^2	ax^4, ax^7
ax^3	$\varnothing$
ax^4	ax^1, ax^2, ax^5, ax^7
ax^5	ax^1, ax^4, ax^6, ax^7
ax^6	ax^1, ax^4, ax^5, ax^7
ax^7	ax^1, ax^2, ax^4, ax^5

Let us turn to Figure c. Similarly, we can get that $J_{(\delta(F),\delta(G))} = 0.9090, J_{(\delta(G),\delta(I))} = 0.8$. According to the transitivity of the equivalence relation, $\delta(F)$and $\delta(I)$should be equivalent, too. However, $J_{(\delta(F),\delta(I))} = 0.7272$. Hence, this is contradictory. By contrasting Figure b and Figure c, we can clearly notice that the distance between sample F and sample I is larger than the distance between sample F and sample H. The distance beyond caused that $\delta(F)$and $\delta(I)$are not equivalent. But the equivalence between $\delta(G)$and $\delta(I)$is not been influent. Hence there is a contradiction in the transitivity of the equivalence relation. If this situation is shown in a concrete form, as shown in Figure d. From Figure d, all the samples in the interval $[D, E)$ and interval $(H, I]$ would have problems in transitivity of equivalence relations.

On the one hand, the above problems can be attributed to the fact that the neighborhood similarity defined by the neighborhood algebraic similarity does not take into account the change of distance between samples, that is, the change of geometric structure. On the other hand, the variable precision neighborhood equivalence relation defined above puts the equivalence relation too relaxed and requires to be constrained.

Definition 2. For any $x^1, x^2 \in U, B \subseteq C$, let $DF^B(x^1, x^2)$be the Minkowsky distance of sample x^1 and sample x^2 in feature space B and δbe the neighborhood radius. The neighborhood geometric similarity between sample x^1 and sample x^2 can be defined as follows.

$$NGS(x^1, x^2) = 1 - \frac{DF^B(x^1, x^2)}{2\delta}. \tag{14}$$

Property 2. *For any $x^1, x^2 \in U, B \subseteq C$, let $DF^B(x^1, x^2)$ be the Minkowsky distance of sample x^1 and sample x^2 in feature space B and δ be the neighborhood radius.*

(1) *If $\delta^B(x^1) \cap \delta^B(x^2) = \emptyset$, namely $DF^B(x^1, x^2) \geqslant 2\delta, NGS(x^1, x^2) = 0$;*
(2) *If $\delta^B(x^1) \cap \delta^B(x^2) \neq \emptyset$, namely $DF^B(x^1, x^2) < 2\delta, NGS(x^1, x^2) > 0$, and the larger $DF^B(x^1, x^2)$ is, the fewer $NGS(x^1, x^2)$ is;*
(3) *If $x^1 = x^2$, namely $DF^B(x^1, x^2) = 0, NGS(x^1, x^2) = 1$;*

Proof.

(1) While $\delta^B(x^1) \cap \delta^B(x^2) = \emptyset$means that the two neighboring granules have no intersection and the distance between the two elements is larger than 2δ, at which point the geometric similarity is 0.
(2) While $\delta^B(x^1) \cap \delta^B(x^2) \neq \emptyset$means that the two neighboring granules have intersection and the distance between the two elements is fewer than 2δ, at which point the geometric similarity is larger than 0.The larger $DF^B(x^1, x^2)$ is, the fewer $NGS(x^1, x^2)$ is.
(3) While $x^1 = x^2$, according to the formula (1), we have $DF^B(x^1, x^2) = 0$, so $NGS(x^1, x^2) = 1$

□

Property 2 represents the relationship between the values of the geometric similarity of the neighborhood we defined in three different cases. (1) When there is no intersection of two neighboring grains, the geometric similarity of the neighborhood is 0 at this time. (2) When two neighboring granules have intersection, the geometric similarity of these two neighboring granules is inversely proportional to the distance between them, the closer the distance the greater the similarity. (3) When the two neighbourhood granules are equal is also the time when the similarity is maximum, then the similarity is 1.

Property 2 mainly reflects the values of the neighborhood geometric similarity we defined in three different cases, which is helpful for us to judge whether the geometric similarity meets the definition of similarity. According to Property 2, we can know that the geometric similarity of the neighborhood has a value range of [0,1], and the maximum value is 1 in the case of overlapping two neighborhood granules. The magnitude of the geometric similarity of the neighborhood decreases as the distance between the two neighborhood granules increases, which is in line with our design expectation.

Neighborhood geometric similarity is mainly calculated by the distance between two samples, and the distance is then normalized by using 2δ. Because if two neighborhoods in geometric space have an intersecting area, the distance between their neighborhood centers must be fewer than 2δ. When there is no intersecting area, the distance between their neighborhood centers must be larger than 2δ. So 2δ is a good measure of whether two neighborhoods have an area of intersection. Therefore, we use 2δ to normalize the distance between two neighborhoods. Since the intersecting area of two neighborhoods is inversely proportional to the sample distance, we use 1 minus the normalized sample distance to represent the neighborhood geometric similarity of two samples.

As mentioned before, the neighborhood similarity defined by the neighborhood algebraic similarity does not take into account the change of geometric structure. To tackle this problem, we combine the neighborhood geometric similarity and the neighborhood algebraic similarity, the two similarities are multiplied and the result is considered as the new neighborhood similarity.

Definition 3. For any $x^1, x^2 \in U, B \subseteq C$, let $DF^B(x^1, x^2)$ be the Minkowsky distance of sample x^1 and sample x^2 in feature space B and δ be the neighborhood radius. Four new neighborhood similarities can be defined as follows.

1. By multiplying the Jaccard similarity and NGS(NGS*Jaccard), a new neighborhood similarity can be defined as follows.

$$NGJS^B(x^1, x^2) = \left(1 - \frac{DF^B(x^1, x^2)}{2\delta}\right) * \frac{\left|\delta^B(x^1) \cap \delta^B(x^2)\right|}{\left|\delta^B(x^1) \cup \delta^B(x^2)\right|}. \tag{15}$$

2. By multiplying the Optimistic similarity and NGS(NGS*Optimistic), a new neighborhood similarity can be defined as follows.

$$NGOS^B(x^1, x^2) = \left(1 - \frac{DF^B(x^1, x^2)}{2\delta}\right) * \frac{\left|\delta^B(x^1) \cap \delta^B(x^2)\right|}{\min\left(\left|\delta^B(x^1)\right|, \left|\delta^B(x^2)\right|\right)}. \tag{16}$$

3. By multiplying the Pessimistic similarity and NGS(NGS*Pessimistic), a new neighborhood similarity can be defined as follows.

$$NGPS^B(x^1,x^2)=\left(1-\frac{DF^B(x^1,x^2)}{2\delta}\right)*\frac{\left|\delta^B(x^1)\cap\delta^B(x^2)\right|}{\max\left(\left|\delta^B(x^1)\right|,\left|\delta^B(x^2)\right|\right)}. \tag{17}$$

4. By multiplying the Average similarity and NGS(NGS*Average), a new neighborhood similarity can be defined as follows.

$$NGAS^B(x^1,x^2)=\left(1-\frac{DF^B(x^1,x^2)}{2\delta}\right)*\frac{\left|\delta^B(x^1)\cap\delta^B(x^2)\right|}{average\left(\left|\delta^B(x^1)\right|,\left|\delta^B(x^2)\right|\right)}. \tag{18}$$

In this paper, we use $NS^B(x^1,x^2)$to on behalf of the four neighborhood similarities.

Property 3. *For any $x^1,x^2\in U,B\subseteq C$, let $DF^B(x^1,x^2)$be the Minkowsky distance of sample x^1 and sample x^2 in feature space B and δbe the neighborhood radius.*

(1) $NGJS^B(x^1,x^2)\in[0,1],NGOS^B(x^1,x^2)\in[0,1],NGPS^B(x^1,x^2)\in[0,1],NGAS^B(x^1,x^2)\in[0,1]$;
(2) $NGJS^B(x^1,x^2)\leqslant NGPS^B(x^1,x^2)\leqslant NGAS^B(x^1,x^2)\leqslant NGOS^B(x^1,x^2)$;
(3) $x^1=x^2\Longleftrightarrow DF^B(x^1,x^2)=0\Longleftrightarrow NGJS^B(x^1,x^2)=NGOS^B(x^1,x^2)=NGPS^B(x^1,x^2)=NGAS^B(x^1,x^2)=1$.

Proof. It can be proved in a similar way as Property 1.

Property 3 mainly reflects whether the four novel neighborhood similarity measures conform to the basic properties of similarity measures and the relationship between them. From Property 3, we can know that the four novel similarity measures take values in the range of [0,1], and their similarity is 1 when the two samples are equal, so they meet the definition of similarity. And these four similarity measures have a size relationship, so that it is also convenient to observe the impact of different sizes of similarity measures on the accuracy rate when comparing the accuracy rate. □

3.2. Variable precision neighborhood rough set model induced by neighborhood similarity

In this subsection, we propose a novel variable precision neighborhood rough set model induced by neighborhood similarity and construct the decision tree algorithm by the degree of attribute dependence.

Definition 4. Let $NDS=\langle U,C\cup D,\delta\rangle$ be a neighborhood decision system, where δ represents the neighborhood radius and β represents the variable precision threshold. For any $B\subseteq C$ and any tuple $(x^1,x^2)\in U\times U$, if the following equation is satisfied

$$VNER^B_{(\delta,\beta)}=\left\{(x^1,x^2)\in U\times U\middle|NS^B(x^1,x^2)\geqslant\beta\right\}, \tag{19}$$

we say the tuple (x,y) satisfies the variable precision neighborhood equivalence relation$VNER^B_{(\delta,\beta)}$.

Specially, when $\delta=0$, the neighborhood granule degenerate into the classical granule, and when $\beta=0$, the neighborhood equivalence relation degenerates into the classical equivalence relation.

Definition 5. Let δ be the neighborhood radius, β be the precision threshold, for any $B\subseteq C$, the variable precision neighborhood equivalence class derived from variable precision neighborhood equivalence relation can be defined as follows.

$$U/VNER^B_{(\delta,\beta)}=\left\{[x]_{VNER^B_{(\delta,\beta)}}|x\in U\right\}=\left\{X^1_{(\delta,\beta)},\ldots,X^i_{(\delta,\beta)}\right\}, \tag{20}$$

where $X^1_{(\delta,\beta)},\ldots,X^i_{(\delta,\beta)}\subseteq U$.

Theorem 1. *Given a decision table $\langle U,A\rangle$, where δ is the neighborhood radius, and β is the variable precision threshold,*

(1) *For any $x^i\in U$, if $\delta^1\leqslant\delta^2$, then we have $\delta^1(x^1)\leqslant\delta^2(x^1)$;*
(2) *If $\beta^1\leqslant\beta^2$, then we have $U/VNER^B_{(\delta,\beta^2)}\subseteq U/VNER^B_{(\delta,\beta^1)}$;*
(3) *If $B^1\subseteq B^2\subseteq C$, then we have $[x]_{VNER^{B^1}_{(\delta,\beta)}}\subseteq[x]_{VNER^{B^2}_{(\delta,\beta)}}$.*

Proof. (1) and (2) are obviously according to Definition. (3) Since $B^1 \subseteq B^2$, it means that $B^2$is rougher classified than B^1. Hence $[x]_{VNER^{B^1}_{(\delta,\beta)}} \subseteq [x]_{VNER^{B^2}_{(\delta,\beta)}}$. □

Theorem 1 focuses on some basic theories of equivalence granule structure. According to Theorem 1 we can know that the same sample may also contain more elements in the case of larger neighborhood radius. In the situation where the threshold of variable precision is larger, the equivalence relation becomes more relaxed, which leads to the equivalence granule containing more elements. In the case of rougher division, the equivalence granules will also be rougher, resulting in the equivalence granules containing more elements.

Definition 6. Let δ be the neighborhood radius, and β be the variable precision threshold. Let $\left\{Y^1, Y^2, \ldots, Y^N\right\}$ be the Equivalent division induced by U/D, for any $B \subseteq C$, the lower approximation of D can be defined as follows.

$$\underline{VNER^B_{(\delta,\beta)}}D = \bigcup_{j=1}^{N} \underline{VNER^B_{(\delta,\beta)}}Y^j, \tag{21}$$

where

$$\underline{VNER^B_{(\delta,\beta)}}Y^j = \bigcup \left\{ X^i_{(\delta,\beta)} \middle| X^i_{(\delta,\beta)} \subseteq Y^j, X^i_{(\delta,\beta)} \in U/VNER^B_{(\delta,\beta)} \right\}, i = 1,2,\ldots,M; j = 1,2,\ldots,N$$

In the formula, $\underline{VNER^B_{(\delta,\beta)}}D$ is called the positive region of decision D, denoted as $POS^B_{(\delta,\beta)}D$.

Theorem 2. *Let $NDS = \langle U, C \cup D, \delta \rangle$ be a neighborhood decision system, if $B^1 \subseteq B^2 \subseteq C$, then we have $POS^{B^1}_{(\delta,\beta)}X \subseteq POS^{B^2}_{(\delta,\beta)}X$.*

Proof. Since $B^1 \subseteq B^2$, it means that $[x]_{VNER^{B^1}_{(\delta,\beta)}} \subseteq [x]_{VNER^{B^2}_{(\delta,\beta)}}$. Hence we have $POS^{B^1}_{(\delta,\beta)}X \subseteq POS^{B^2}_{(\delta,\beta)}X$. □

Definition 7. Let δ be the neighborhood radius, and β be the variable precision threshold. Let $\left\{Y^1, Y^2, \ldots, Y^N\right\}$ be the Equivalent division induced by U/D, for any $B \subseteq C$, the upper approximation of D can be defined as follows.

$$\overline{VNER^B_{(\delta,\beta)}}D = \bigcup_{j=1}^{N} \overline{VNER^B_{(\delta,\beta)}}Y^j, \tag{22}$$

where

$$\overline{VNER^B_{(\delta,\beta)}}Y^j = \bigcup \left\{ X^i_{(\delta,\beta)} \middle| X^i_{(\delta,\beta)} \in U/VNER^B_{(\delta,\beta)} : X^i_{(\delta,\beta)} \cap Y^j \neq \emptyset \right\}, i = 1,2,\ldots,M; j = 1,2,\ldots,N$$

Definition 8. Let δ be the neighborhood radius, and β be the variable precision threshold. For any $B \subseteq C$, the degree of attribute dependence of decision attribute set D upon B can be defined as follows.

$$r_{(\delta,\beta)}(D,B) = \frac{\left|\underline{VNER^B_{(\delta,\beta)}}D\right|}{|U|}. \tag{23}$$

Proposition 1. *Let δ be the neighborhood radius, and β be the variable precision threshold. For any $B \subseteq C$, let $U/VNER^B_{(\delta,\beta)}$ be the variable precision neighborhood granule, we have*

(1) *When $\beta = 1, r_{(\delta,\beta)}\left(D, U/VNER^B_{(\delta,\beta)}\right) = r_{(\delta,\beta)}\left(D, U/NER^B_{(\delta)}\right)$;*
(2) *When $\delta = 0$ and $\beta = 1$, we have $r_{(\delta,\beta)}\left(D, U/VNER^B_{(\delta,\beta)}\right) = r\left(D, U/R^B\right)$.*

Proof. When $\beta = 1$, the variable precision neighborhood granule degenerates into the rigorous equivalent neighborhood granule. So the variable precision neighborhood attribute dependence degenerates into the neighborhood attribute dependence. When $\delta = 0$ and $\beta = 1$, the variable precision neighborhood granule degenerates into the Pawlak equivalent granule. So the variable precision neighborhood attribute dependence degenerates into the Pawlak attribute dependence. □

Table 5
The details of datasets.

	Dataset	Sample	Condition attribute	Decision attribute
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	*ENB*2012	768	9	6
13	*magic*	2501	10	2
14	*winequality*	4899	11	7

Table 6
The accuracy(%).

	Dataset	ID3	CART	C4.5	NGS∗Jaccard	NGS∗Aver	NGS∗Opti	NGS∗Pessi
1	*Crayo*	0.8556	0.8445	0.8	1	1	1	0.9778
2	*Iris*	0.95	0.9333	0.9833	0.9533	0.9567	0.9833	0.97
3	*wine*	0.9889	0.9556	0.9667	0.9222	0.9306	0.9444	0.9615
4	*plrx*	0.879	0.9	0.9	0.9221	0.9342	0.9342	0.9474
5	*wpbc*	0.9	0.925	0.935	0.9125	0.9	0.925	0.9
6	*Sheesegmentation*	0.8	0.8476	0.8333	0.881	0.9286	0.9524	0.9524
7	*seeds*	0.881	0.8571	0.881	0.9048	0.9167	0.9048	0.9405
8	*glass*	0.7159	0.6931	0.7841	0.9205	0.9205	0.9205	0.9318
9	*heart*	0.9148	0.9222	0.9407	0.9537	0.9537	0.9167	0.9537
10	*ecoli*	0.9191	0.9117	0.8897	0.9485	0.9485	0.9118	0.9412
11	*ILPD*	0.8093	0.7754	0.8305	0.9322	0.9322	0.911	0.9237
12	*ENB*2012	0.2039	0.2143	0.213	0.9188	0.9188	0.899	0.9058
13	*magic*	0.8801	0.873	0.8608	0.911	0.926	0.908	0.917
14	*winequality*	0.6638	0.6699	0.6699	0.9311	0.9484	0.9217	0.9355
15	*Average*	0.8115	0.8086	0.8206	0.9344	0.9357	0.9396	0.9421

Theorem 3. *Let δ be the neighborhood radius, and β be the variable precision threshold, if $B^1 \subseteq B^2 \subseteq C$, then we have* $r_{(\delta,\beta)}\left(D, U/VNER^{B^1}_{(\delta,\beta)}\right) \subseteq r_{(\delta,\beta)}\left(D, U/VNER^{B^2}_{(\delta,\beta)}\right)$.

Proof. According to Theorem 2, we can obtain that $POS^{B^1}_{(\delta,\beta)}X \subseteq POS^{B^2}_{(\delta,\beta)}X$. Hence, $r_{(\delta,\beta)}\left(D, U/VNER^{B^1}_{(\delta,\beta)}\right) \leqslant r_{(\delta,\beta)}\left(D, U/VNER^{B^2}_{(\delta,\beta)}\right)$. □

3.3. Algorithm design

As mentioned before, many classical decision tree algorithms can only deal with the discrete data. The continuous data should be preprocessed by data discretization. In order to avoid the loss of information structure due to data discretization, we have to find new measure. The algorithm constructed in this paper uses the degree of attribute dependency from the variable precision neighbourhood rough set model as a new measure and adaptively generates neighbourhood radius based on the data itself, making the algorithm itself more robust.

When we do neighborhood granulation on the decision table, since different samples have different values under different condition attributes. When the differences of values are very large, it is easy to lead to unreasonable neighborhood granulation if we use the same neighborhood radius to all condition attributes. Under such conditions, we need a neighborhood radius that adaptively generates as the samples change in the range of values in the condition attributes. Because the standard deviation is one of the most frequently adopted quantitative forms to reflect the degree of the dispersion of a set of data, it is suitable for the neighborhood radius. Moreover, to reduce the influence of noise, the size of the neighborhood radius can be adjusted by the classification accuracy coefficient *Lambda*. The classification accuracy coefficient *Lambda* is generally set to 2 to 4, and the neighborhood of the attribute is obtained together.

$$\delta_{R^1} = \frac{sta\left(R^1\right)}{Lambda}, \tag{24}$$

where $sta\left(R^1\right) = \sqrt{\frac{1}{H}\sum_{i=1}^{H}(x^i - \overline{x})}$ is the standard deviation of the samples in attribute R^1.

The time complexity of Algorithm 1 is mainly concentrated in Step 1. Neighborhood granulation requires a granulation operation for each element to form a neighborhood and the time complexity is $O\left(|C| * |U|^2\right)$. The next equivalent division is the same as neighborhood granulation, which requires a division operation over all elements, and the time complexity is $O\left(|C| * |U|^2\right)$. Calculating the attribute dependency is simply a computation for each attribute, and the time complexity is $O(|C|)$. So the time complexity of Algorithm 1 is $O\left(|C| * |U|^2\right)$.

Algorithm 1: Decision Tree Algorithm Based on Variable Precision Neighborhood Similarity

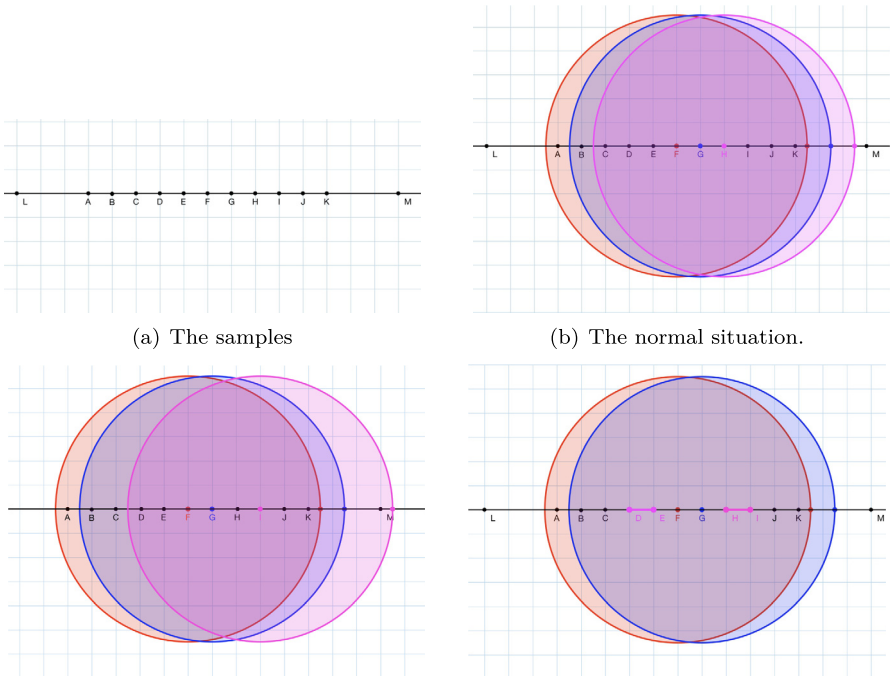
Input: A decision table $NDS = \langle U, C \cup D, \delta \rangle$
Output: A variable precision neighborhood decision tree
1. **for** each $B \subseteq C$ **do**

Compute $r_{(\delta,\beta)}(D,B) = \frac{\left|\underline{VNER^B_{(\delta,\beta)}}D\right|}{|U|}$;

end for
2. Select the attribute set $B' \in B$ with the max value of attribute dependency $B' = arg_{\forall B \subset C} \max r_{(\delta,\beta)}(D,B)$, randomly select an optimal attribute as a split node;
3. Equivalent granules $U/VNER^B_{(\delta,\beta)} = \left\{X_B^1, X_B^2, \ldots, X_B^n\right\}$ formed by variable precision neighborhood equivalence relations divide the selected attributes into n decision subtables, each granule is a subtable and a branch;

For each decision subtable **do**
4. If the decision label of all samples are the same, use the decision label to create a leaf node;
5. If $B' = \emptyset$, use the maximum number of the decision label to create a leaf node in parent node
6. If the value of the samples are different with the same decision label, go to step 1.

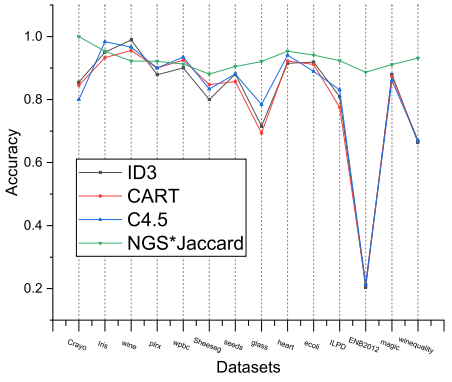
End For
7. Output a variable precision neighborhood decision tree.



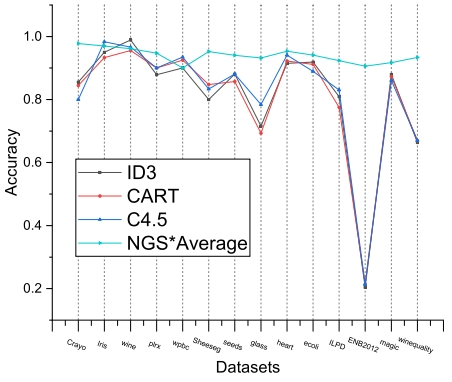
(a) The samples (b) The normal situation.

(c) The situation that satisfies algebraic relationship but does not satisfy transitivity. (d) problem situation normalization.

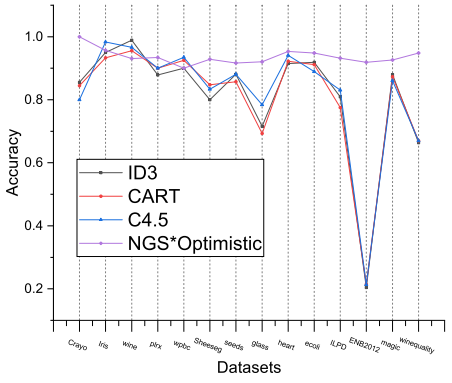
Fig. 1. Example 1 and 2.



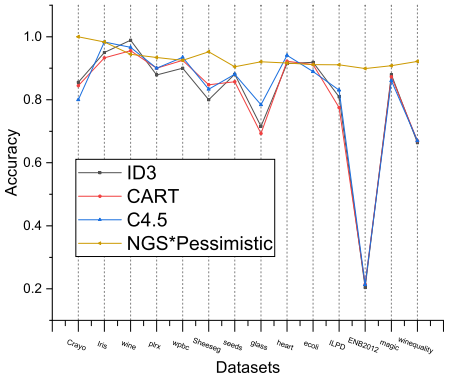
(a) NGS*Jaccard



(b) NGS*Average

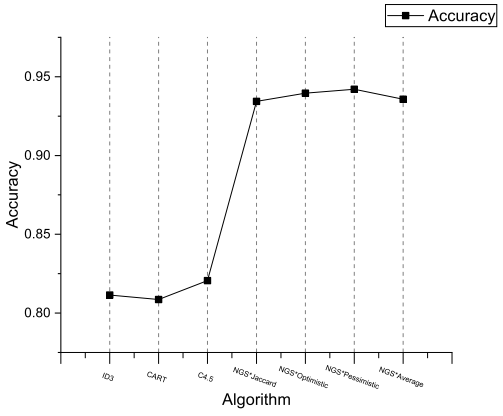


(c) NGS*Optimistic

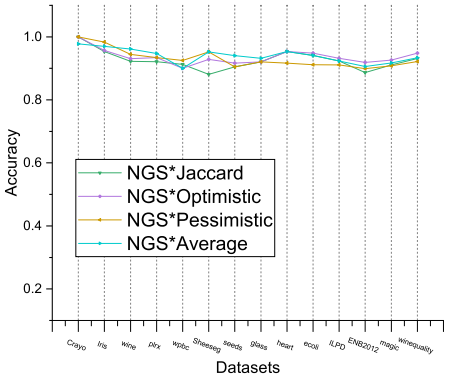


(d) NGS*Pessimistic

Fig. 2. The accuracy comparison of three classical algorithms with new algorithms.



(a) The averagment of the accuracy.

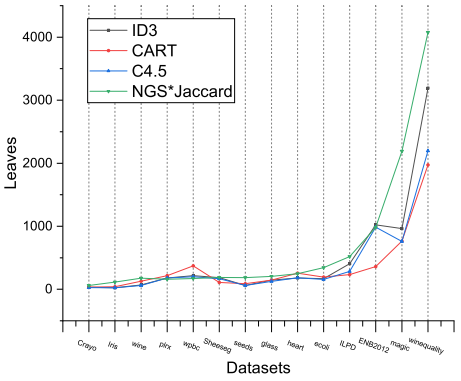


(b) The accuracy comparison of four algorithm.

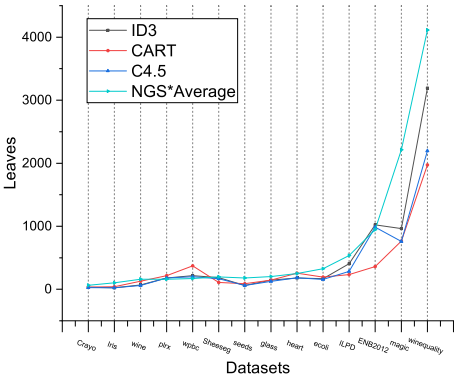
Fig. 3. The averagment of the accuracy and the comparison of four algorithm.

Table 7
The number of leaves.

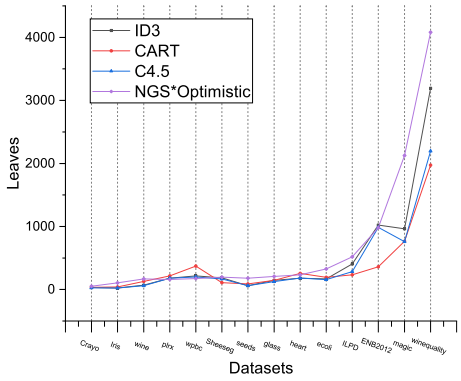
	Dataset	ID3	CART	C45	NGS∗Jaccard	NGS∗Aver	NGS∗Opti	NGS∗Pessi
1	*Crayo*	32	35	27	60	53	75	62
2	*Iris*	23	40	24	113	106	110	104
3	*wine*	69	128	59	176	168	171	158
4	*plrx*	178	216	178	159	159	159	159
5	*wpbc*	215	372	200	172	173	169	170
6	*Sheesegmentation*	184	108	169	188	197	190	195
7	*seeds*	61	89	59	186	179	184	178
8	*glass*	145	145	124	205	207	204	202
9	*heart*	178	255	182	247	233	253	251
10	*ecoli*	169	192	157	344	326	336	326
11	*ILPD*	405	234	285	519	520	517	537
12	*ENB*2012	1023	361	986	984	980	961	947
13	*magic*	963	764	758	2190	2126	2115	2213
14	*winequality*	3194	1973	2196	4076	4080	4079	4112
15	*Average*	488	350	385	686	679	679	686



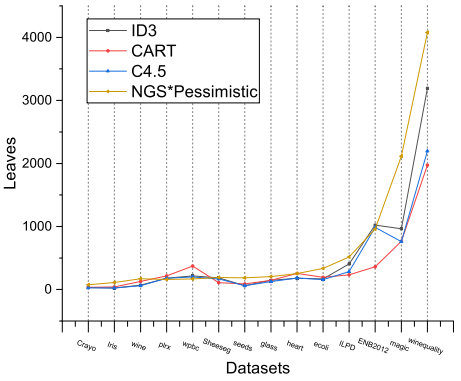
(a) NGS*Jaccard



(b) NGS*Average

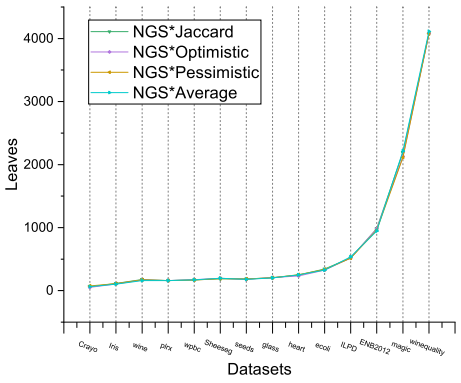


(c) NGS*Optimistic

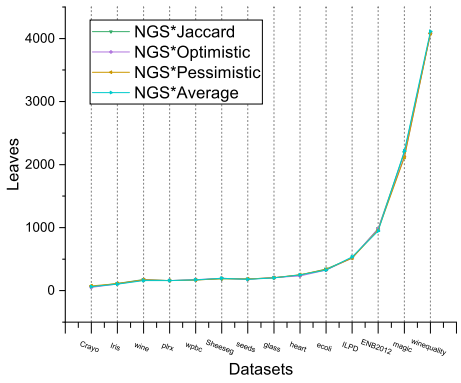


(d) NGS*Pessimistic

Fig. 4. The number of leaves comparison of three classical algorithms with new algorithms.



(a) The averagment of number of leaves.



(b) The number of leaves of four algorithm.

Fig. 5. The averagment of number of leaves and the number of leaves of four algorithm.

4. Experimental analysis

In this section, we experimentally demonstrate the effectiveness of the improved decision tree algorithm. We selected 14 datasets from the UCI Machine Learning Repository, as shown in Table 5. For the comparison with the classical decision tree algorithms, the dataset is discretized by equal distance partition in advance, and the specific partition interval is 5. To estimate the effectiveness of the decision tree, we use the 10-fold cross-validation method to test and verify the accuracy and the number of leaves. The classification accuracy coefficient Lambda is 2, and the precision threshold $\beta = 0.8$. All of the algorithms are run in MATLAB2021b and the hardware environment with Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz 2.59 GHz and 16.0 GB RAM.

From Table 6, in terms of accuracy, the algorithm proposed in this paper increases by an average of more than 10 % compared with the classical algorithms. The most obvious in the 14th dataset, the classical algorithms have only 20 % accuracy, while the proposed algorithm is basically above 90% (Fig. 1).

Fig. 2 shows the specific comparison of the accuracy among the algorithm introduced in this paper and three classical algorithms. From the figure, we can see that in most datasets, the algorithm proposed in this paper outperforms the classical algorithms. Fig. 3 shows the average accuracy of all algorithms and the comparison results of four separate algorithms.

From Table 7, it is clearly that the number of leaves generated by the algorithm introduced in this paper is more than that of the classical algorithms, indicating that the degree of fitting of our algorithm is higher than that of the classical algorithms, but it does not reach the degree of over-fitting, and the number of leaves generated is also within the acceptable range.

Fig. 4 shows the specific comparison of the number of leaves among the algorithm introduced in this paper and the three classical algorithms. From the figures, it is clearly that in most datasets, the algorithm proposed in this paper are larger than the classical algorithms. Fig. 5 shows the average number of the leaves of all algorithms and the comparison results of four algorithms. The results demonstrate that the algorithm introduced in this paper is highly feasible in general.

5. Conclusion

In this paper, we first define the neighborhood geometric similarity by considering the geometric structure in the neighborhood system and define a new neighborhood similarity by combining it with the neighborhood algebraic similarity. Secondly, we construct a novel variable precision neighborhood rough set model by using the new neighborhood similarity. The novel model solves the problem of possible contradiction in the transferability of equivalence relations in the old model that only considered the neighborhood algebraic similarity. Then, we construct a decision tree algorithm using the attribute dependency in the new model as the partition measure. Finally, we present experimental comparisons showing the computational efficiency of the proposed methods and the experimental results show that the proposed methods are effective and efficient.

CRediT authorship contribution statement

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

Declaration of 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.

Acknowledgment

The research is supported by the National Natural Science Foundation of China under Grant Nos. 62166001, 61976158, Natural Science Foundation of Jiangxi Province, China, under Grant No. 20202BAB202010, Graduate Innovation Funding Program of Gannan Normal University, China under Grant No. YCX22A025, the Project of Science and Technology of Education Department of Jiangxi Province, under Grant No.GJJ211407.

References

[1] Duoqian Miao, Jue Wang, Rough sets based approach for multivariate decision tree construction, Chin. J. Software 8(6) (1997) (in Chinese with English Abstract).
[2] Leo Breimann, Jerome H. Friedman, Richard A. Olshen, Charles J. Stone, Classification and regression trees, Pacific Grove, Wadsworth, 1984.
[3] J.R. Quinlan et al, Induction of decision trees. 1986 Kluwer Academic Publisher,[Retrieved on Mar. 15, 2012], Retrieved from the Internet 26 (1986) 82–106.
[4] Steven L. Salzberg, C4. 5: Programs for machine learning by j. ross quinlan, Morgan Kaufmann Publishers, Inc., 1993, 1994.
[5] Yingyue Chen, Yumin Chen, Feature subset selection based on variable precision neighborhood rough sets, Int. J. Comput. Intell. Syst. 14 (1) (2021) 572–581.
[6] John Y. Ching, Andrew K.C. Wong, Keith C.C. Chan, Class-dependent discretization for inductive learning from continuous and mixed-mode data, IEEE Trans. Pattern Anal. Mach. Intell. 17 (7) (1995) 641–651.
[7] Zdzisław Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[8] Xiuyi Jia, Lin Shang, Bing Zhou, Yiyu Yao, Generalized attribute reduct in rough set theory, Knowl.-Based Syst. 91 (2016) 204–218.
[9] Yu.hua. Qian, Jiye Liang, Witold Pedrycz, Chuangyin Dang, Positive approximation: an accelerator for attribute reduction in rough set theory, Artif. Intell. 174 (9–10) (2010) 597–618.
[10] Yumin Chen, Duoqian Miao, Ruizhi Wang, A rough set approach to feature selection based on ant colony optimization, Pattern Recogn. Lett. 31 (3) (2010) 226–233.
[11] Hu. Qinghua, Witold Pedrycz, Daren Yu, Jun Lang, Selecting discrete and continuous features based on neighborhood decision error minimization, IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) 40 (1) (2009) 137–150.
[12] Jianming Zhan, Qiumei Wang, Certain types of soft coverings based rough sets with applications, Int. J. Mach. Learn. Cybern. 10 (5) (2019) 1065–1076.
[13] Fannia Pacheco, Mariela Cerrada, René-Vinicio Sánchez, Diego Cabrera, Chuan Li, José Valente de Oliveira, Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery, Expert Syst. Appl. 71 (2017) 69–86.
[14] Hu. Qinghua, Yu. Daren, Zongxia Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2) (2008) 866–876.
[15] Qinghua Hu, Daren Yu, Jinfu Liu, Congxin Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.
[16] Can Gao, Zhicheng Wang, Jie Zhou, Three-way approximate reduct based on information-theoretic measure, Int. J. Approx. Reason. 142 (2022) 324–337.
[17] Can Gao, Yong Li, Jie Zhou, Witold Pedrycz, Zhihui Lai, Jun Wan, Lu. Jianglin, Global structure-guided neighborhood preserving embedding for dimensionality reduction, Int. J. Mach. Learn. Cybern. 13 (7) (2022) 2013–2032.
[18] Can Gao, Jie Zhou, Duoqian Miao, Xiaodong Yue, Jun Wan, Granular-conditional-entropy-based attribute reduction for partially labeled data with proxy labels, Inf. Sci. 580 (2021) 111–128.
[19] Feng Jiang, Yuefei Sui, Cungen Cao, An incremental decision tree algorithm based on rough sets and its application in intrusion detection, Artif. Intell. Rev. 40 (4) (2013) 517–530.
[20] Mikhail Moshkov, On the depth of decision trees with hypotheses, Entropy 24 (1) (2022) 116.
[21] Rong Wang, Zunren Liu, Ji Jun, Decision tree algorithm based on attribute importance, Chin. Comput. Sci. 44 (Z11) (2017) 129–132 (in Chinese with English Abstract).
[22] Wojciech Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1) (1993) 39–59.
[23] Mingfeng Ye, The datamining algorithm on knowledge dependence, in: 2018 International Conference on Smart Grid and Electrical Automation (ICSGEA), IEEE, 2018, pp. 234–236.
[24] Xin Xie, Zhang Xianyong, Wang Wuanye, Tang Pengfei, Neighborhood decision tree construction algorithm based on variable precision neighborhood equivalent granules, Chin. J. Comput. Appl. (2021) (in Chinese with English Abstract).
[25] Linhai Cheng, Rough set model of interval rough number under similarity relation (Master's thesis), Guangxi University, 2020.
[26] Jun Ye, Multicriteria decision-making method using the dice similarity measure between expected intervals of trapezoidal fuzzy numbers, J. Decis. Syst. 21 (4) (2012) 307–317.
[27] Ziqi Yan, Qiong Wu, Meng Ren, Jiqiang Liu, Shaowu Liu, Shuo Qiu, Locally private jaccard similarity estimation, Concurrency and Computation: Practice and Experience 31 (24) (2019) e4889.
[28] Meng Qi, Yunfan Gu, Qiong Wang, Internet financial risk management and control based on improved rough set algorithm, J. Comput. Appl. Math. 384 (2021) 113179.
[29] Eric C.C. Tsang, Bingzhen Sun, Weimin Ma, General relation-based variable precision rough fuzzy set, Int. J. Mach. Learn. Cybern. 8 (3) (2017) 891–901.
[30] Juan Li, Yabin Shao, Xiaoding Qi, On variable-precision-based rough set approach to incomplete interval-valued fuzzy information systems and its applications, J. Intell. Fuzzy Syst. 40 (1) (2021) 463–475.
[31] Shuyin Xia, Zhao Zhang, Wenhua Li, Guoyin Wang, Elisabeth Giem, Zizhong Chen, Gbnrs: a novel rough set algorithm for fast adaptive attribute reduction in classification, IEEE Trans. Knowl. Data Eng. (2020).