Soft Computing
https://doi.org/10.1007/s00500-024-09835-4

FOUNDATION, ALGEBRAIC, AND ANALYTICAL METHODS IN SOFT COMPUTING

An improved decision tree algorithm based on hierarchical neighborhood dependence

Jianying Lai[1] · Caihui Liu[1] · Bowen Lin[1] · Duoqian Miao[2]

Accepted: 15 March 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Neighborhood rough sets (NRS) is widely used in various fields with good adaptability, and its related information measurement plays an important role in uncertainty analysis. In the existing research, although the three-layer granular structure of neighborhood decision system (NDS) has been proposed, there are not many related research on it, especially in the uncertainty measurement. Therefore, this paper firstly deeply discusses the expression of neighborhood dependence in the three-layer granularity structure of neighborhood decision system, as well as its relationship in three levels and some related properties. Secondly, considering the influence of the fixed neighborhood radius(Nr) on the neighborhood model, we define the adaptive neighborhood radius by using the standard deviation and the neighborhood mean. Finally, we use the neighborhood dependence at the macro top level as the split node measurement function to construct the decision tree. Our experimental results show the reliability of the algorithm.

Keywords Neighborhood rough sets · Decision tree · Three-level analysis · Hierarchical neighborhood dependence

1 Introduction

Rough sets (Pawlak 1982) is a mathematical tool proposed by Professor Z.Pawlak of Warsaw University of Technology in Poland to deal with uncertain and fuzzy data. It has shown a strong ability to describe uncertain or inaccurate knowledge without prior knowledge. Therefore, rough sets are widely used in data mining (Luo et al. 2018; Saha et al. 2019; Zhang et al. 2014), machine learning and knowledge discovery (Qian et al. 2018; Eissa et al. 2016), artificial intelligence (Bishop and Nasrabadi 2006; Dai et al. 2012) and many other fields. However, the classical rough set theory with limitation of processing discrete data. For continuous data, the strategy of discretization is adopted in data preprocessing, generally, which may cause information loss and calculation error. With the increment of data scale, this shortcoming will be continuously amplified, and the classical rough set cannot be well applied to big data. Based on the above shortcomings, scholars who study rough sets have generalized the traditional theory and proposed more adaptable models (Hu et al. 2008; Yao 1998).

NRS is one of the extended models of classical rough set, which is proposed by Hu et al. (2008). In the NRS model, the strict equivalence relation in rough set is replaced by neighborhood relation, which reduces the information loss caused by data discretization and more suitable for the current situation of data explosion. As an applicable model, the selection of neighborhood radius is a key issue in neighborhood model. The size of the radius of the neighborhood affects the thickness of the neighborhood granule, and whether the radius is large or small will have a certain impact on the entire model. However, in most of the existing models, the Nr implements an empirical setting approach. Hence, in order to decrease the influence, many scholars have developed diverse radius constructions. Qu et al. (2023) proposed a novel neighbor-

✉ Caihui Liu
liucaihui@gnnu.edu.cn

Jianying Lai
1421631021@qq.com

Bowen Lin
641017193@qq.com

Duoqian Miao
dqmiao@tongji.edu.cn

[1] Department of Mathematics and Computer Science, Gannan Normal University, Ganzhou 341000, Jiangxi, China

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

Published online: 20 July 2024

hood radius via calculating the distance between boundary samples in different decision classes, and then averaging all the distances. This radius can reflect samples in the same class have the same granularity space, and different samples have different granularity space. Sun et al. (2022) combined the classification accuracy of the selected attributes, the average correlation degree of the reduced attribute set, and the reduction rate of the classification results to construct a new adaptive Nr function to avoid artificially selecting the optimal neighborhood radius. In the study of NRS-based multi-label feature selection, Liu et al. (2023) defined the adaptive radius based on the neighborhood relationship constructed by using the distribution information of similar instances, and obtained good results in the experiment. The granular ball NRS proposed by Xia et al. (2020, 2019) enables each object to adaptively generate a neighborhood, which tackles the problems of the empirical setting of neighborhood radius. Although several scholars have developed studies about the radius, the adaptive model is still worth researching due to the complex data environment in real applications.

Granular computing (Yao 2016, 2001; Zhang and Miao 2014), as a structured method of hierarchical computing and granular processing, provides several multi-level and multi-granularityblue (Bakkouri et al. 2022; Bakkouri and Afdel 2023) research methods, and brings some novel strategies to uncertainty analysis and knowledge reasoning. For instance, Zhang and Miao (2017) first introduced the three-layer granular structure of decision table, and hierarchicalized the classical rough set. Chen et al. (2019) introduced analytic hierarchy process into neighborhood rough sets and proposed the three-layer granular structure for neighborhood system. The introduction of hierarchical analysis makes the structure of the neighborhood system more clear, which provides conditions for the improvement of the neighborhood system. In addition, we can further analyze the characteristics of different levels of granular structure and the correlation among them. Zhang et al. focused on the three-layer granular structure of NRs, and deeply studied the uncertainty measures such as neighborhood conditional entropy, neighborhood complementary entropy and neighborhood mutual information at different levels (Gou and Zhang 2021; Liao et al. 2021; Mu et al. 2019; Tang et al. 2020; Zhang and Yao 2022; Zhang et al. 2021). This shows that the three-layer granular structure of the neighborhood system still has significance to explore.

The decision tree algorithm (Breiman 2017; Quinlan 1996; Liu et al. 2022, 2023) is one of the classic classification (Safavian and Landgrebe 1991) methods and it has been widely used owing to its high intelligibility. However, it cannot handle continuous data directly without the strategy of discretization. Generally, the discretization may leads to the loss of important information contained in the data, which has a certain impact on the classification performance. Neighborhood rough sets as a theoretical tool that can be directly applied to continuous data, has certain advantages in data processing. However, there is a lack of research about the improved decision tree algorithm based on NRS. Xie et al. (2022); Xin et al. (2022) proposed a neighborhood decision tree algorithm by combining it with decision tree algorithm, and obtained good experimental results. Based on this research, this paper further studies the improved method. We find that according to the corresponding equivalence partitioning method in the above research, using neighborhood information gain as the metric function of node selection with the disadvantage of the tendency to depend on the attributes with more equivalence classes, and the attributes with more equivalence classes are not necessarily important attributes, resulting in poor generalization performance. In addition, it requires a large number of logarithmic operations, which increases the difficulty of calculation. In this paper, the neighborhood dependence degree is used as the node measurement function, which can not only solve the problems mentioned above, but also fully consider the relationship between conditional attributes and decision attributes. The greater the dependence of decision attributes on conditional attributes, the more specific information they contain. Therefore, the conditional attribute with the highest degree of dependence is selected as the current node. This method not only speeds up the growth of the decision tree, but also can obtain a well-structured decision tree and convenient for mining good rule information.

Based on the above research, this paper focuses on the three-layer granular structure of the NDS, studying the expression of the neighborhood dependence at the macro-high level, meso-middle level, and micro-bottom level and the relationship of it between three levels, and discuss some related properties of it. At the same time, focusing on the selection of Nr, an adaptive Nr is defined to reduce the influence on the neighborhood model. Subsequently, an improved decision tree algorithm is constructed by using the neighborhood dependence at the macro-high level as the partition measure. Our main contributions of this paper are as follows:

- Completely explored the expression of dependence under different granularity levels.
- A new adaptive Nr is defined to reduce the influence of radius selection on the model.
- A novel decision tree algorithm is designed by using the neighborhood dependency as the node metric function.

The remaining structure of this paper is as follows. Section 2 briefly reviews some related uncertainty measures in decision tree algorithms and the neighborhood rough set model. Section 3 analyzes the three-layer granular structure of neighborhood decision system and constructs the neighborhood dependence on three-layer granular structure. Then,

we define the adaptive neighborhood radius and design a more robust decision tree algorithm that is combined with the neighborhood rough set. In Sect. 4, we verify the effectiveness of the proposed algorithm on public datasets. Finally, we make a summary in Sect.5.

2 Preliminaries

In this section, we first review some knowledge of decision tree. As a classical classification algorithm that has been developed for decades, its powerful classification function is still worthy of our in-depth study. A review of the original classical decision tree algorithm can provide a knowledge base for subsequent research. Next, we will introduce some well-known decision tree metric functions. Throughout the paper, U is called universe to represent a non-empty finite set.

2.1 Decision tree

Decision tree algorithm, as a classical classification algorithm that has been developed for decades, is still worthy of a deepen study. The powerful uncertainty measures of the algorithm have also been well applied to different fields and achieved good performance. In this subsection, the partition measure in ID3 and CART algorithms are briefly reviewed.

Definition 1 Let $DT = (U, A \cup D)$ be a decision table, where A is the conditional feature set and D is the decision feature set. Given the decision classification $U/D = \{D_1, D_2, ..., D_k\}$, the information entropy $Ie(D)$ of U/D is defined as follows:

$$Ie(D) = -\sum_{r=1}^{k} \frac{|D_r|}{|U|} \log \frac{|D_r|}{|U|} \tag{1}$$

Given any attribute subset $C \subseteq A$, let $U/C = \{C_1, C_2, ..., C_m\}$, the conditional entropy $Ie(D|C)$ of D with respect to C is defined as:

$$Ie(D|C) = \sum_{j=1}^{m} \frac{|C_j|}{|U|} Ie(C_j) \tag{2}$$

where $1 \le j \le m$, $Ie(C_j) = -\sum_{i=1}^{k} \frac{|C_j \cap D_r|}{|C_j|} \log \frac{|C_j \cap D_r|}{|C_j|}$. Then, the information gain of C is

$$Gain(C) = Ie(D) - Ie(D|C) \tag{3}$$

Definition 2 Let $DT = (U, A \cup D)$ be a decision table, and given the decision classification $U/D = \{D_1, D_2, ..., D_k\}$, the gini index $Gini(D)$ of U/D is defined as:

$$\begin{aligned} Gini(D) &= \sum_{r=1}^{k} \frac{|D_r|}{|U|} \left(1 - \frac{|D_r|}{|U|}\right) \\ &= 1 - \sum_{r=1}^{k} \left(\frac{|D_r|}{|U|}\right)^2 \end{aligned} \tag{4}$$

Given any $C \subseteq A$, let $U/C = \{C_1, C_2, ..., C_m\}$, the gini index $Gini(D, C)$ of D with respect to conditional attribute C is defined:

$$Gini(D, C) = \sum_{j=1}^{m} \frac{|C_j|}{|U|} Gini(C_j) \tag{5}$$

where $1 \le j \le m$ and $Gini(C_j) = 1 - \sum_{r=1}^{k} \left(\frac{|C_j \cap D_r|}{|C_j|}\right)^2$.

2.2 Neighborhood rough set model

NRS is proposed by Hu et al. (2008), which is the generalization of classical rough set. Compared with classical rough sets, NRS use neighborhood relation to replace the equivalence relation, so as to divide the universe and generate equivalence classes. The proposed of neighborhood relation solves the problem that the classical rough set with the difficulty of processing continuous data, so it is widely used in various fields. Here, we depict some essential knowledges about it.

Definition 3 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, where U represents a non-empty finite set with samples $\{a_1, a_2, a_3, ..., a_n\}$ and is called universe.$C = \{c_1, c_2, c_3, ...c_n\}$ is a set of conditional feature variable, and D represents the decision feature. Given a sample $a_i \in U$ and $B \subseteq C$, the neighborhood $\delta_B(a_i)$ of a_i with respect to B is defined as

$$\delta_B(a_i) = \{a_j \in U | \Delta^B(a_i, a_j) \le \delta\} \tag{6}$$

where Δ is a distance function, as for $\forall a_1, a_2, a_3 \in U$, Δ always satisfies,

(1) $\Delta(a_1, a_2) \ge 0$,$\Delta(a_1, a_2) = 0$ only when $a_1 = a_2$;
(2) $\Delta(a_1, a_2) = \Delta(a_2, a_1)$;
(3) $\Delta(a_1, a_3) \le \Delta(a_1, a_2) + \Delta(a_2, a_3)$.

Given any two samples $\forall a_1, a_2 \in U$ with respect to $B \subseteq C$,where $B = \{c_1, c_2, c_3, ...c_n\}$, $d_B(a_1, a_2)$ represents the distance between sample a_1 and a_2 on attribute set B, then the Minkowsky distance is expressed as

$$d_B(a_1, a_2) = \left(\sum_{k=1}^{n} |f(a_1, c_i) - f(a_2, c_i)|^p\right)^{1/p} \tag{7}$$

where $f(a_1, c_i)$ is the value of sample a_1 on attribute c_i, $1 \leq i \leq n$.

For the distance function above, if $p = 1$, it is called Manhattan distance, if $p = 2$, named Euclidean distance, if $p = \infty$, named Chebyshev distance. Generally, the Manhattan distance is used in the NRS.

Definition 4 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, for any set $X \subseteq U$ and feature subset $B \subseteq C$, the upper approximation and the lower approximation of set X on B are defined as follows,

$$\overline{N}_B(X) = \{a_i \in U | \delta(a_i) \cap X \neq \phi\} \tag{8}$$

$$\underline{N}_B(X) = \{a_i \in U | \delta(a_i) \subseteq X\}. \tag{9}$$

At the same time, we denotes the boundary region of set X on B as $B_N(X) = \overline{N}_B(X) - \underline{N}_B(X)$.

Definition 5 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, let $U/D = \{D_1, D_2, ..., D_n\}$ be the equivalence classes(Es) on U. For any $B \subseteq C$, we have

$$POS_B(D) = \underline{N}_B(D) \tag{10}$$

where $\underline{N}_B(D) = \bigcup_{i=1}^{n} \underline{N}_B D_i$ and $\underline{N}_B D_i = \{a_i | \delta_B(a_i) \subseteq X, a_i \in U\}$.

Definition 6 Let $NDs = (U, C \cup D, V, f, \delta)$ be a NDS, the dependency of D with respect to B is expressed as

$$\gamma_B(D) = \frac{|POS_B(D)|}{|U|} \tag{11}$$

3 Hierarchical neighborhood dependence degree based on three-level granular structures

In this section, firstly, we briefly review the three-layer granular structure of NDS, then study the uncertainty measure of neighborhood dependence in the three-layer granular structure in detail and propose the hierarchical neighborhood dependence. Finally, an example is given to illustrate.

3.1 Three-layer granular structure in neighborhood decision system

In recent years, with the study of three levels of thinking, it has been gradually developed and applied. Zhang and Miao (2017) introduced three-layer thinking into rough set theory and proposed a three-layer granularity structure. With the deepening of research, it has been successfully applied to classical decision systems. Subsequently, Zhang et al. (2022) proposed a three-layer granular structure for the neighborhood decision system and analyzed it. This subsection we will briefly review and research it.

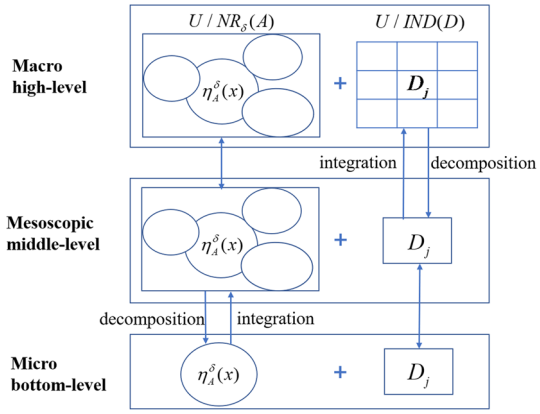

Fig. 1 Hierarchical relationship and structure of three-layer granular structure (Zhang and Miao 2017)

Figure 1 shows the three-layer granular structure of neighborhood decision system in detail. Next, we will discuss in more detail from the information contained in the figure to understand the system more clearly.

Let $NDs = (U, C \cup D, V, f, \delta)$ be a NDS, $U/IND(D) = \{D_1, D_2, ...D_n\}$ represents the decision classification, for any feature subset $A \subseteq C$,$U/NR^\delta(A)$ represents the Es of U on A, $\eta_A^\delta(x)$ is the δ neighborhood of sample x on feature subset A.We can get the following conclusions.

We can obviously find the following characteristics by observing the whole decision system top-down from the macro high-level.

(1) The macro top-level can be decomposed into $k = 2^A - \{\emptyset\}$ meso middle-level, and the middle-level can be decomposed into $|U|$ micro bottom-level.
(2) The macro top-level can be decomposed into $k * |U|$ micro bottom-level.

Conversely, we can find the following characteristics by looking at the whole decision system bottom-up.

(1) The related $|U|$ micro bottom-level can form meso middle-level, and k middle-level can compose a macro high-level.
(2) All $k * |U|$ micro bottom-level can compose macro high-level.

Table 1 gives a general description of the three-layer granular structure of the NDS. Neighborhood granules at the micro-bottom mainly focus on the interaction with a decision class, describing a relationship between granules. We can observe that the universe is divided into n equivalence

Table 1 Three-level granules of neighborhood decision system (Zhang and Miao 2017)

Structure	Composition	Granular scale/level	Simple name
1	$U/IND(A)$,$U/IND(D)$	Macro/top	Macro-top
2	$U/IND(A)$,D_j	Meso/middle	Meso-middle
3	$\eta_A^{\delta}(x)$,D_j	Micro/bottom	Micro-bottom

classes and interacts with a decision equivalence class under the attribute subset A at the meso-middle level. The n equivalence classes are related to the micro neighborhood granules. The macro-level involves dependency relationship between n equivalence classes and all decision equivalence classes.

3.2 Hierarchical neighborhood dependence degree

In the theory of NRS, neighborhood dependence is an important uncertainty measure, which plays an important role in knowledge reasoning and data mining. However, in previous studies, the measurement at the macro level is mostly used. With the introduction of the three-layer granular structure, the expressions of neighborhood dependence between different levels become clearer. In this subsection, we propose a hierarchical neighborhood dependence based on the three-layer granular structure and discuss it in detail.

Definition 7 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, for any feature subset $A \subseteq C$, $\delta_A(x)$ is the neighborhood granule of sample x on the attribute subset A, and D_j is a decision class in U/IND. At the micro-level, the upper and lower approximations of D_j regarding $\delta_A(x)$ are defined as

$$\begin{aligned} \underline{N}(D_j|\delta_A(x)) &= \{x|\delta_A(x) \subseteq D_j, x \in U\} \\ \overline{N}(D_j|\delta_A(x)) &= \{x|\delta_A(x) \cap D_j \neq \emptyset, x \in U\} \end{aligned} \tag{12}$$

Then, we can get the neighborhood dependence and the positive region of D_j on $\delta_A(x)$ is defined as

$$POS(D_j|\delta_A(x)) = \underline{N}(D_j|\delta_A(x)) \tag{13}$$

$$\gamma(D_j|\delta_A(x)) = \frac{POS(D_j|\delta_A(x))}{|U|} \tag{14}$$

According to the relevant definitions at the bottom-level, the hierarchical neighborhood dependence is related to the neighborhood granule $\delta_A(x)$ and the class D_j. Let $\delta_A(x)$ be contained in the decision class D_j, then the $\delta_A(x)$ belongs to the lower approximation $\underline{N}(D_j|\delta_A(x))$, which represents the positive region. Hence, it can be obtained that the POS$(D_j|\delta_A(x))$ is dependent on the certain decision class D_j, and the recognition ability of existing knowledge to certain decision class $D_j$is reflected by the measure $\gamma(D_j|\delta_A(x))$.

Definition 8 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, $U/NR(A)$ is the equivalence classes of the universe U on the feature subset A, denoted as ∂_A, and D_j is a decision class in U/IND. At the meso-level, the upper and lower approximations of D_j regarding ∂_A are defined as

$$\begin{aligned} \underline{N}(D_j|\partial_A) &= \cup_1^{|U|} \underline{N}(D_j|\delta_A(x)) \\ \overline{N}(D_j|\partial_A) &= \cup_1^{|U|} \overline{N}(D_j|\delta_A(x)) \end{aligned} \tag{15}$$

Then, we can get the neighborhood dependence and the positive region of D_j on ∂_A are defined as

$$POS(D_j|\partial_A) = \underline{N}(D_j|\partial_A) \tag{16}$$

$$\gamma(D_j|\partial_A) = \frac{POS(D_j|\partial_A)}{|U|} \tag{17}$$

According to the structure and related definitions at the middle-level, the hierarchical neighborhood dependence involves the interaction of the Es $U/NR^{\delta}(A)$ and the class D_j. The equivalence class $U/NR^{\delta}(A)$ is formed of $|U|$ equivalence granules $\delta_A(x)$, similarly, $U/NR^{\delta}(A)$ can be decomposed into $|U|$ equivalent granules $\delta_A(x)$. The union of $|U|$ $\underline{N}(D_j|\delta_A(x))$ forms the lower approximation of the meso-level with fixed class D_j, and then the positive region $POS(D_j|\partial_A)$ is obtained. Therefore, the hierarchical neighborhood dependence at the meso-level is $\gamma(D_j|\partial_A) = \frac{POS(D_j|\partial_A)}{|U|}$.

Theorem 1 *Hierarchical neighborhood dependence degree shows the characteristics of hierarchical decomposition / integration,as shown in Fig. 1.*

$$\begin{aligned} \gamma(D_j|\partial_A) &= \frac{POS(D_j|\partial_A)}{|U|} \\ &= \frac{\underline{N}(D_j|\partial_A)}{|U|} \\ &= \frac{\cup_1^{|U|} \underline{N}(D_j|\delta_A(x))}{|U|} \\ &= \frac{\cup_1^{|U|} \{x|\delta_A(x) \subseteq D_j, x \in U\}}{|U|} \end{aligned} \tag{18}$$

It can be seen from Theorem 1, the hierarchical neighborhood dependence depends on the positive region $POS(D_j|\partial_A)$ from the perspective of the middle-level, and the

positive region is equivalent to the lower approximation $\underline{N}(D_j|\partial_A)$. ∂_A is composed of $|U|$ equivalent granules, so $\underline{N}(D_j|\partial_A)$ is decomposed into $|U|$ lower approximations of neighborhood granules with fixed decision class D_j placed at the bottom-level. at the bottom-level, the lower approximation is determined by the neighborhood granule contained in the fixed decision class D_j. The decomposition of this level shows the relationship between the middle-level and the bottom-level of the hierarchical neighborhood dependence. The hierarchical neighborhood dependence at the middle-level, it can be decomposed downward according to the equivalence class ∂_A, and also can be integrated upward from the bottom-level.

Proposition 1 *Given $\langle U, A, NR\rangle$ and two neighborhood radii δ_1 and δ_2,for any attribute $C \subseteq A$, if $\delta_1 \leq \delta_2$, we have*

(1) $\forall e_i \in U, NR_1 \subseteq NR_2, \delta_1(e_i) \subseteq \delta_2(e_i)$;
(2) $\forall E \subseteq U : \underline{NR_1}E \supseteq \underline{NR_2}E; \overline{NR_1}E \supseteq \overline{NR_2}E$.

where NR_1 and NR_2 represent the neighborhood relations induced by δ_1 and δ_2, respectively.

***Proof* 1** If $\delta_1 \leq \delta_2$,explicitly, we have $\delta_1(e_i) \subseteq \delta_2(e_i)$ and then $NR_1 \subseteq NR_2$. Suppose $\delta_2(e_i) \subseteq E$, we can get $\delta_1(e_i) \subseteq E$. Therefore, if we let $e_i \in \underline{NR_2}E$,then $e_i \in \underline{NR_1}E$. However, e_i is not necessary in $\underline{NR_2}E$ if we have $e_i \in \underline{NR_1}E$. Hence, $\underline{NR_1}E \supseteq \underline{NR_2}E$.And we can get $\overline{NR_1}E \supseteq \overline{NR_2}E$ similarly.

Proposition 1 shows that the selection of neighborhood radius affects the size of neighborhood granules, thus affecting the lower approximation. The smaller neighborhood radius has a larger lower approximation, and the larger neighborhood radius has a smaller lower approximation. □

Proposition 2 *Given $\langle U, A \cup D, NR, \delta\rangle$,for any attribute $C_1 \subseteq C_2 \subseteq A$, we have*

(1) If $C_1 \subseteq C_2$,$[e]_{NER_\delta^{C_1}} \subseteq [e]_{NER_\delta^{C_2}}$
(2) If $C_1 \subseteq C_2, POS_\delta^{C_1}E \subseteq POs_\delta^{C_2}E$

***Proof* 2** If $C_1 \subseteq C_2$,We will say that C_2 has a rougher classification than C_2,so we have $[e]_{NER_\delta^{C_1}} \subseteq [e]_{NER_\delta^{C_2}}$.Therefore, $POS_\delta^{C_1}E \subseteq POS_\delta^{C_2}E$. □

Definition 9 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, $U/NR(A)$ is the equivalence class of the universe U on the attribute subset A, denoted as ∂_A, and let $U/IND(D) = \{D_1, D_2, ..., D_m\}$ be the decision equivalence classes on U,denoted as ∂_D. At the macro-level, the upper and lower approximations of ∂_D regarding ∂_A are defined as

$$\begin{aligned}\underline{N}(\partial_D|\partial_A) &= \cup_1^m \underline{N}(D_j|\partial_A)\\ \overline{N}(\partial_D|\partial_A) &= \cup_1^m \overline{N}(D_j|\partial_A)\end{aligned} \tag{19}$$

Then, we can get the neighborhood dependence and the positive region of D_j on ∂_A are defined as

$$POS(\partial_D|\partial_A) = \underline{N}(\partial_D|\partial_A) \tag{20}$$

$$\gamma(\partial_D|\partial_A) = \frac{POS(\partial_D|\partial_A)}{|U|} \tag{21}$$

At the macro top-level, the hierarchical neighborhood dependence is related to the number of decision equivalence classes $U/IND(D)$. m is the number of decision equivalence classes, and the union of m $\underline{N}(D_j|\partial_A)$ at the middle-level form $\underline{N}(\partial_D|\partial_A)$. The positive region is determined by the lower approximation $\underline{N}(\partial_D|\partial_A)$, so we can get $\gamma_\delta(\partial_D|\partial_A) = \frac{POS(\partial_D|\partial_A)}{|U|}$.

Theorem 2 *Hierarchical neighborhood dependence degree at the macro-level is related to the number of decision equivalence classes.*

$$\begin{aligned}\gamma(\partial_D|\partial_A) &= \frac{POS(\partial_D|\partial_A)}{|U|}\\ &= \frac{\underline{N}(\partial_D|\partial_A)}{|U|}\\ &= \frac{\cup_1^m \underline{N}(D_j|\partial_A)}{|U|}\end{aligned} \tag{22}$$

Theorem 2 reflects the relationship between high-level and middle-level of hierarchical neighborhood dependence. From the perspective of macro high-level, the hierarchical neighborhood dependence is related to the number of decision equivalence classes. The hierarchical neighborhood dependence at the macro high-level, it can be decomposed into the lower approximation $\underline{N}(\partial_D|\partial_A)$ of m decision equivalence classes with respect to ∂_A at the middle-level and also can be integrated upward from the middle-level.

Proposition 3 *Given $\langle U, A \cup D, NR, \delta\rangle$, for any attribute $C \subseteq A$, let U/NER_δ^C be neighborhood granular,if $\delta = 0$ we have $\gamma_\delta(D, U/NER_\delta^C) = \gamma(D, U/R^C)$.*

***Proof* 3** Obviously, if if $\delta = 0$, neighborhood rough sets will degenerate into classical Pawlak rough sets. Hence $\gamma_\delta(D, U/NER_\delta^C) = \gamma(D, U/R^C)$.

Proposition 3 shows that NRs is the generalization of classical rough set. When the neighborhood radius $\delta = 0$, the neighborhood relation degenerate into classical equivalence relation. Hence, neighborhood rough set degenerate into Pawlak rough set. □

Proposition 4 *Given $\langle U, A \cup D, NR, \delta\rangle$,for any attribute $C_1 \subseteq C_2 \subseteq A$, let U/NER_δ^C be neighborhood granular,we have $\gamma_\delta(D, U/NER_\delta^{C_1}) \leq \gamma_\delta(D, U/NER_\delta^{C_2})$.*

Table 2 A given decision Table

U	C_1	C_2	C_3	C_4	D
e_1	0.5	0.8	1	0.9	Y
e_2	0.8	0.1	0.5	0	N
e_3	0.5	0.4	0.6	0.3	Y
e_4	0.2	0.7	0.9	0.7	Y
e_5	0.1	0.3	0.3	0.6	Y
e_6	0.9	0.1	0.7	0.6	N

***Proof* 4** According to Proposition 2, we can get that $POS_\delta^{C_1} E \subseteq POS_\delta^{C_2} E$,hence $\gamma_\delta(D, U/NER_\delta^{C_1}) \leq \gamma_\delta(D, U/NER_\delta^{C_2})$.

Proposition 4 describes the monotonicity of neighborhood dependence degree. The increase of features in the feature subset does not reduce the dependence, showing a tendency to increase monotonically with the increase of features. Then, we use Example 1 to describe the relationship of the neighborhood dependence between different levels in the three-layer granular structure. □

Example 1 Given a dataset table shown as Table 2. We use an example to explore the close relationship between the three-layer granular structure. According to the definition of the three-level granular structure, we first give the various measurement results of the neighborhood granules under the micro-bottom layer as shown in Table 3, and then calculate the relevant measurement results according to the different definitions of the other two levels. Herein, we give $\delta = 0.3$.

In the meso middle-level, according to the above definition, then we can calculate the lower and upper approximations of D_1 and D_2 regarding C_1. The results are as follows.

(1) $\underline{N}(D_1|\partial_{C_1}) = \cup_1^6 \underline{N}(D_1|\delta_{C_1}(x)) = \{e_4, e_5\}$,
(2) $\overline{N}(D_1|\partial_{C_1}) = \cup_1^6 \overline{N}(D_1|\delta_{C_1}(x)) = \{e_1, e_2, e_3, e_6\}$,
(3) $\underline{N}(D_2|\partial_{C_1}) = \cup_1^6 \underline{N}(D_2|\delta_{C_1}(x)) = \{e_6\}$,
(4) $\overline{N}(D_2|\partial_{C_1}) = \cup_1^6 \overline{N}(D_2|\delta_{C_1}(x)) = \{e_1, e_2, e_3, e_4, e_5\}$.

Then, we can get their dependency degree as follows,

(1) $\gamma(D_1|\partial_{C_1}) = \frac{POS(D_1|\partial_{C_1})}{|U|} = \frac{1}{3}$,

(2) $\gamma(D_2|\partial_{C_1}) = \frac{POS(D_2|\partial_{C_1})}{|U|} = \frac{1}{6}$.

The calculation results at the macro-level and the middle level are closely related to it. Through the above calculation, we can get

(1) $\underline{N}(D|\partial_{C_1}) = \cup_1^2 \underline{N}(D|\partial_{C_1}) = \{e_4, e_5, e_6\}$,
(2) $\overline{N}(D|\partial_{C_1}) = \cup_1^2 \overline{N}(D|\partial_{C_1}) = \{e_1, e_2, e_3, e_4, e_5, e_6\}$.

Regarding the calculation of dependence degree, we have $\gamma(D|\partial_{C_1}) = \frac{POS(D|\partial_{C_1})}{|U|} = \frac{1}{2}$

3.3 Adaptive neighborhood and algorithm design

Definition 10 Let $NDS = (U, C \cup D, V, f, \delta)$ be a NDS, and δ is neighborhood radii parameter, for any $B \subseteq C$, the neighborhood equivalence relation on B is denoted as follows:

$$NER_B^\delta = \{(x, y) \in U \times U | \eta_B^\delta(x) = \eta_B^\delta(y)\} \tag{23}$$

where $\delta \in [0, 1]$, and $\eta_B^\delta(*)$ represents the δ-neighborhood of sample $*$ on B. The neighborhood equivalence division on U induced by the neighborhood equivalence relation NER_B^δ is expressed as follows:

$$U/NER_B^\delta = \{[y]_{NER_B^\delta} | y \in U\} = \{Y_1^B, Y_2^B, ..., Y_n^B\} \tag{24}$$

where $Y_r^B = \{y_i \in U | \eta_B^\delta(y_i) = \eta_B^\delta(y_j)\}; y_j \in Y_r^B (r = 1, 2, ..., n)$.

According to the above expression of neighborhood equivalence, we can conclude that when the neighborhoods of two samples are exactly equal, they are considered to be equal. This equivalence relation can be used as the dividing standard of decision tree algorithm.

The selection of Nr will affect the applicability of the whole neighborhood model, due to the fact that the size of the Nr determines the roughness of certain neighborhood granules. Therefore, how to select an appropriate Nr is a key issue. It is known that the distribution of attribute values can excellently reflect the characteristics of numerical attributes.

Table 3 Correlation measurement results at the micro-bottom

U	Equivalent class of C_1	$\underline{N}(D_1)$	$\overline{N}(D_1)$	$\gamma(D_1)$	$\underline{N}(D_2)$	$\overline{N}(D_2)$	$\gamma(D_2)$
e_1	$\{e_1,e_2,e_3,e_4\}$	Ø	$\{e_1\}$	0	Ø	$\{e_1\}$	0
e_2	$\{e_1,e_2,e_3,e_6\}$	Ø	$\{e_2\}$	0	Ø	$\{e_2\}$	0
e_3	$\{e_1,e_2,e_3,e_4\}$	Ø	$\{e_3\}$	0	Ø	$\{e_3\}$	0
e_4	$\{e_1,e_3,e_4,e_5\}$	$\{e_4\}$	Ø	1/6	Ø	$\{e_4\}$	0
e_5	$\{e_4,e_5\}$	$\{e_5\}$	Ø	1/6	Ø	$\{e_5\}$	0
e_6	$\{e_2,e_6\}$	Ø	$\{e_6\}$	0	$\{e_6\}$	Ø	1/6

Hence, this paper constructs the adaptive radius mainly based on standard deviation and mean value.

Definition 11 In $NDS = (U, C \cup D, V, f, \alpha)$, for the condition feature set $C = \{c_1, c_2, ..., c_n\}$, its adaptive neighborhood radius set is expressed as

$$\alpha = \{\alpha_1, \alpha_2, ..., \alpha_n\} \tag{25}$$

where $\alpha_i = \frac{std(c_i)}{mean(c_i)}$ and $1 \le i \le n$.

$std(c_i)$ is the standard deviation of the values of all objects under condition attribute c_i, and $mean(c_i)$ is the mean of the values of all objects under this attribute. Obviously, $\frac{std(c_i)}{mean(c_i)} \in (0, 1)$

The standard deviation reflects the average fluctuation size of the data on the average value, showing a normal distribution. The mean can reflect the concentration trend of a set of data. The combination of standard deviation and mean can make the obtained neighborhood radius not appear too large or too small as the normal distribution median, and can better constrain the radius size. The neighborhood radius under each attribute can be adjusted independently by using this method to select the radius. The effectiveness of the adaptive radius will be reflected in the construction of the decision tree.

After determining the value of the neighborhood radius, we give the following algorithm by combining NRS and decision tree.It is known that the NRs can handle the continuous data without the strategy of discretization. Hence, for the decision tree algorithm with the problem that cannot fit continuous data directly, the combination of neighborhood rough sets is one of the methods to solve this shortcoming, which broadens the applicability of decision tree algorithm. Moreover, the rough set technology with more complex output results and the decision tree technology with simpler output results are integrated with each other to achieve complementary advantages. The experimental algorithm is designed as follows:

The time complexity involved in this algorithm is mainly generated by the calculation of step 1, the uncertainty measure $\gamma(D, c)$. Firstly, we need to granulate all the samples in the universe to generate neighborhood granules, the time complexity is $O(M * N^2)$, M is the number of attributes, and N denotes the number of samples in the universe. Secondly, the calculation of attribute dependency is relatively simple, and its time complexity is $O(M)$. Finally, the equivalence partition of the decision tree needs to find the equivalence class of each attribute, and its time complexity is $O(M * N^2)$. Therefore, the time complexity of the algorithm is $O(M * N^2)$.

Table 4 A new decision algorithm

Algorithm: A decision tree algorithm combined with NRs
Input: $NDs = (U, A \cup D, V, f, \alpha)$,$\alpha$ is adaptive neighborhood radius, attribute subset $C \subseteq A$
Output:A new decision tree based on NRs
1: **For** each C **do**
Calculate the $\gamma(D, c)$ placed at the top of the macro level;
End For
2: Determine the max value $\max_{c \in C \subseteq A} \gamma(D, c)$ and randomly select an optimal attribute from arg $\max_{c \in C \subseteq A} \gamma(D, c)$ as the current splitting attribute;
3: According to the above selected optimal attributes, calculate the neighborhood equivalence division $U/NER_C^{\alpha} = \{X_1^C, X_2^C, ..., X_n^C\}$, where each granule structure (i.e., X_i^C, $1 \le i \le n$) denotes a branch;
For each branch **do**
4: If all samples with the same decision class, create leaf nodes under this decision class;
5: If $B = \emptyset$, use the decision class with the largest number of samples to create leaf node;
6: Otherwise, turn to step 1;
End For
7: Return A new decision tree based on NRs.

Table 5 The describe of dataset

No	Dataset	*Samples*	Conditional attribute	Classes
1	Iris	150	4	3
2	Wine	178	13	3
3	Glass	214	10	6
4	Crayo	90	6	2
5	Ecoil	336	7	7
6	Plrx	182	18	1
7	Wpbc	194	33	2
8	ILPD	583	10	2
9	Heart	270	13	2
10	Seg	210	19	7
11	END2012	768	9	5
12	Segment	2310	19	7

4 Experimental analysis

In this section, we verify the effectiveness of the proposed algorithm. Here, we use 12 public datasets to verify its reliability and evaluate the algorithm from different perspectives.

4.1 Data sets

We select 13 public UCI datasets as the experimental data. The description of these datasets are as as Table 5:

Table 6 The accuracy of algorithms

No	Data	ID3	CART	C4.5	VPNDT	**TD**
1	Crayo	0.6667	0.5556	0.6786	0.8333	**1**
2	Iris	0.6	0.6667	0.6	0.8	**0.9333**
3	Wine	0.8333	0.6806	0.8611	0.7222	**0.9444**
4	Plrx	0.6842	0.6164	0.7368	0.7838	**0.9474**
5	Wpbc	0.8	0.5769	0.7692	0.8462	**0.95**
6	Seg	0.619	0.5833	0.619	0.8571	**0.9524**
7	Glass	0.6364	0.6395	0.6818	0.7907	**0.9545**
8	Heart	0.8148	0.6481	0.8148	0.8333	**0.963**
9	Ecoil	0.7941	0.7822	0.8088	0.7941	**0.9706**
10	ILPD	0.7458	0.7561	0.8136	0.8376	**0.9831**
11	END2012	0.3636	0.3701	0.4156	0.7922	**0.961**
12	Segment	0.8918	0.7792	0.8831	0.8312	**0.9351**
13	Arithmetic Mean	0.7041	0.6379	0.7235	0.8101	**0.9579**

4.2 Experiment setting

This experiment is carried out in the hardware environment of Intel(R) Core(TM) i5-11300H @3.10GHz 3.11 GHz, RAM 16.0GB, and the comparative algorithms include ID3 algorithm, C4.5 algorithm, CART algorithm and VPNDT algorithm. All data sets are continuous data. Since the above algorithms cannot be directly applied to continuous data, we use the approach of equi-distant for discretization. The performance of the effectiveness of the decision tree algorithm is usually reflected by the two indicators of accuracy and leaf number. In this experiment, these two indicators are used to evaluate the algorithm. In order to verify its effectiveness, ten-fold cross-validation is used.

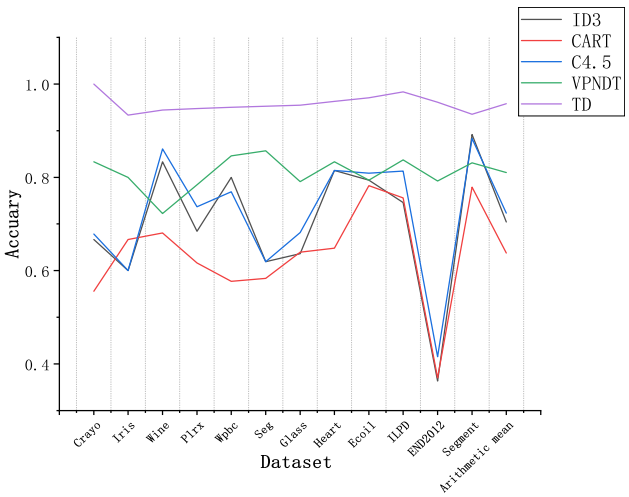

Fig. 2 The accuracy description diagram of four algorithms with TD algorithm

4.3 Experimental results and analysis

Table 6 lists the accuracy of our algorithm, the three classical algorithms and the VPNDT algorithm in the latest literature on 13 datasets. The data show that our algorithm have achieved good performance in classification, and its accuracy is greater than 90%. Compared with the other four algorithms, the improvement of accuracy is considerable. From the perspective of average accuracy, the TD algorithm is more than 30% higher than the classical algorithm.

Figure 2 visualizes the accuracy of the algorithms. It can be seen that the improvement of the proposed algorithm is obvious. In particular, the performances on the dataset Wpbc, END2012, Glass, and Iris are very obvious. The proposed algorithm can not only be applied to continuous data directly, but also has better performance, which shows that our improvement has reference significance. Therefore, the proposed algorithm is effective.

Table 7 describes the number of leaves of four algorithms and TD algorithm. We can see it presented by each algorithm on different data sets is not much different. Compared with the other four algorithms, the number of leaves in this algorithm is slightly lower, and there is no big gap, so our algorithm has good performance. From the average number of leaves, our algorithm is slightly higher within the acceptable range, which indicates that our algorithm is still effective.

Figure 3 is an intuitive representation of the number of leaves of five algorithms on 12 data sets. It can be seen from the figure that the number of leaves of the TD algorithm does not increase dramatically. It maintains the same number as other algorithms, and even has better performance on some datasets. Although there is a higher number of leaves in parts of datasets, this is acceptable. Therefore, the performance of TD algorithm is effective.

Table 7 The leaves of algorithms

No.	Data	ID3	CART	C4.5	VPNDT	**TD**
1	Crayo	34	52	26	51	**41**
2	Iris	16	46	16	108	**57**
3	Wine	76	237	62	161	**153**
4	Plrx	172	233	186	140	**160**
5	Wpbc	207	391	173	153	**171**
6	Seg	174	288	174	187	**200**
7	Glass	149	180	130	184	**188**
8	Heart	176	334	165	222	**225**
9	Ecoil	180	242	151	305	**302**
10	ILPD	525	542	419	462	**526**
11	END2012	1224	787	1227	825	**930**
12	Segment	1479	1842	1187	1754	**1987**
13	Arithmetic Mean	367.67	431.17	326.33	379.33	**411.67**

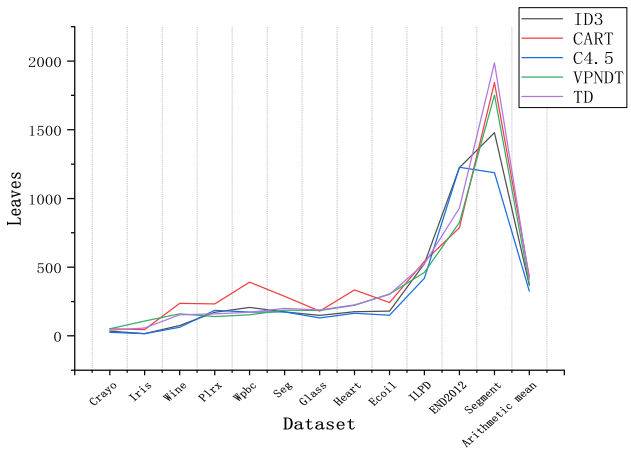

Fig. 3 The leaves description of four algorithms with TD algorithm

5 Conclusions

In this paper, the three-layer granular structure of NDS is studied, and then extended to neighborhood decision system, and the three-layer granular structure under neighborhood decision system is studied. The related properties between different structural levels and their internal relations are discussed, and the uncertainty measure of dependence is deeply studied, and the dependence between different levels is determined. Aiming at the influence of fixed Nr on neighborhood model, we define adaptive neighborhood to improve the model. Then, the neighborhood rough sets are combined with the decision tree algorithm to solve the problem that the decision tree algorithm can not handle continuous data directly, and a new decision tree algorithm is proposed. Focusing on the existing research using neighborhood information gain as the metric function of node selection, which has the tendency of relying on more attributes of equivalence classes and the need for a large number of logarithmic operations, we adopt the method of using neighborhood dependence as the metric function of nodes, which not only solves the above problems, but also fully considers the relationship between conditional attributes and decision attributes. We have verified the method on 12 data sets, and achieved good accuracy. Compared with other existing methods, it shows better classification results. This shows that our improved method is effective.In the future, considering the combination of neighborhood model and decision tree to improve the accuracy of classical algorithms is still worthy of our further research.

Author Contributions Jianying Lai: conceptualization, methodology, writing—original draft, software, validation. Caihui Liu: conceptualization, methodology, writing—review editing, validation, supervision. Bowen Lin: data curation, software. Duoqian Miao: writing—review editing, supervision.

Funding The research is supported by the National Natural Science Foundation of China under Grant Nos. 62166001, 61976158.

Data availability Data will be made available on request.

Declarations

Conflict of 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.

Ethical and informed consent for data used All the data used in the paper are from the UC Irvine Machine Learning Repository.

References

Bakkouri I, Afdel K (2023) MLCA2F: multi-level context attentional feature fusion for COVID-19 lesion segmentation from CT scans. SIViP 17(4):1181–1188

Bakkouri I, Afdel K, Benois-Pineau J, Initiative GCFTASDN (2022) BG-3DM2F: bidirectional gated 3D multi-scale feature fusion for Alzheimer's disease diagnosis. Multimed Tools Appl 81(8):10743–10776

Bishop CM, Nasrabadi NM (2006) Pattern recognition and machine learning no 4, vol 4. Springer, New York. p, p 738

Breiman L (2017) Classification and regression trees. Routledge, New York

Chen Y, Qin N, Li W, Xu F (2019) Granule structures, distances and measures in neighborhood systems. Knowl-Based Syst 165:268–281

Dai J, Xu Q, Wang W, Tian H (2012) Conditional entropy for incomplete decision systems and its application in data mining. Int J Gen Syst 41(7):713–728

Eissa MM, Elmogy M, Hashem M (2016) Rough? Granular computing knowledge discovery models for medical classification. Egypt Inf J 17(3):265–272

Gou H, Zhang X (2021) Compromised multi-granulation rough sets based on an attribute-extension chain. J Intell Fuzzy Syst 40(1):715–732

Hu Q, Yu D, Liu J, Wu C (2008) Neighborhood rough set based heterogeneous feature subset selection. Inf Sci 178(18):3577–3594

Hu Q, Yu D, Xie Z (2008) Neighborhood classifiers. Expert Syst Appl 34(2):866–876

Liao S, Zhang X, Mo Z (2021) Three-level and three-way uncertainty measurements for interval-valued decision systems. Int J Mach Learn Cybern 12:1459–1481

Liu C, Lin B, Lai J, Miao D (2022) An improved decision tree algorithm based on variable precision neighborhood similarity. Inf Sci 615:152–166

Liu J, Lin Y, Du J, Zhang H, Chen Z, Zhang J (2023) ASFS: a novel streaming feature selection for multi-label data based on neighborhood rough set. Appl Intell 53(2):1707–1724

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

Luo C, Li T, Chen H, Fujita H, Yi Z (2018) Incremental rough set approach for hierarchical multicriteria classification. Inf Sci 429:72–87

Mu T, Zhang X, Mo Z (2019) Double-granule conditional-entropies based on three-level granular structures. Entropy 21(7):657

Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11:341–356

Qian Y, Liang X, Wang Q, Liang J, Liu B, Skowron A, Dang C (2018) Local rough set: a solution to rough data analysis in big data. Int J Approx Reason 97:38–63

Quinlan JR (1996) Learning decision tree classifiers. ACM Comput Surv (CSUR) 28(1):71–72

Qu K, Xu J, Han Z et al (2023) Maximum relevance minimum redundancy-based feature selection using rough mutual information in adaptive neighborhood rough sets. Appl Intell 53:17727–17746.

Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 21(3):660–674

Saha I, Sarkar JP, Maulik U (2019) Integrated rough fuzzy clustering for categorical data analysis. Fuzzy Sets Syst 361:1–32

Sun L, Zhang J, Ding W, Xu J (2022) Mixed measure-based feature selection using the Fisher score and neighborhood rough sets. Appl Intell 52(15):17264–17288

Tang L, Zhang X, Mo Z (2020) A weighted complement-entropy system based on tri-level granular structures. Int J Gen Syst 49(8):872–905

Xia S, Liu Y, Ding X, Wang G, Yu H, Luo Y (2019) Granular ball computing classifiers for efficient, scalable and robust learning. Inf Sci 483:136–152

Xia S, Zhang H, Li W, Wang G, Giem E, Chen Z (2020) GBNRS: a novel rough set algorithm for fast adaptive attribute reduction in classification. IEEE Trans Knowl Data Eng 34(3):1231–1242

Xie X, Zhang X, Yang J (2022) Improved ID3 decision tree algorithm induced by neighborhood equivalence relations. Comput Appl Res. 39(01): 102-105+112. (In Chinese with English Abstract)

Xin Xie, Xianyong Zhang, Xuanye Wang, Pengfei Tang (2022) Neighborhood decision tree construction algorithm of variable precision neighborhood equivalent Kernel. J Comput Appl 42(02):382–388 ((**In Chinese with English Abstract**))

Yao YY (1998) Relational interpretations of neighborhood operators and rough set approximation operators. Inf Sci 111(1–4):239–259

Yao YY (2001) Information granulation and rough set approximation. Int J Intell Syst 16(1):87–104

Yao Y (2016) A triarchic theory of granular computing. Granul Comput 1:145–157

Zhang X, Miao D (2014) Quantitative information architecture, granular computing and rough set models in the double-quantitative approximation space of precision and grade. Inf Sci 268:147–168

Zhang X, Miao D (2017) Three-layer granular structures and three-way informational measures of a decision table. Inf Sci 412:67–86

Zhang X, Yao Y (2022) Tri-level attribute reduction in rough set theory. Expert Syst Appl 190:116187

Zhang J, Wong JS, Li T, Pan Y (2014) A comparison of parallel large-scale knowledge acquisition using rough set theory on different MapReduce runtime systems. Int J Approx Reason 55(3):896–907

Zhang X, Gou H, Lv Z, Miao D (2021) Double-quantitative distance measurement and classification learning based on the tri-level granular structure of neighborhood system. Knowl-Based Syst 217:106799

Zhang X, Zhou Y, Tang X, Fan Y (2022) Three-way improved neighborhood entropies based on three-level granular structures. Int J Mach Learn Cybern 13(7):1861–1890

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.