2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

A Decision Tree Based On Related Family

Wenxing Li[1[0009−0001−5528−6961]], Xin Yang[2[0009−0006−9362−0971]], Meihua Liu[3[0009−0004−7210−5722]] and Tian Yang[*[1111-2222-3333-4444]]

Hunan Normal University, Lushan Road 36, Changsha 410081, China
liwenxing1997@qq.com
math_yangtian@126.com

Abstract. Decision trees are widely used supervised learning models known for their simplicity, interpretability, and effectiveness in classification and regression tasks. Feature selection can remove redundant and noisy features, enhancing the generalization and robustness of decision trees. However, due to the high computational cost of existing feature selection methods, it is typically applied only once before classifier training, providing the classifier with dimensionally reduced data. This limits the synergistic effect between feature selection and the construction of split nodes in decision trees. The Related Family is an efficient feature evaluation method proposed by our research team. Its efficiency allows us to use it in the construction of split nodes in decision trees, leading to better splitting criteria. Building on this method, we introduce the Dynamic Related Family Decision Tree (DRFDT), which dynamically selects optimal features for each sample subgroup as the tree grows. Experiments demonstrate that DRFDT outperforms a wide range of classification algorithms across 15 UCI datasets, achieving an average accuracy of 89.30%. This represents significant improvements over classical single-feature decision tree methods (CART: +3.87%), traditional classification algorithms (KNN: +5.71%, SVM: +4.54%), multi-feature split decision tree algorithms (CART-LC: +3.99%, O1: +4.25%), and state-of-the-art decision tree classification algorithms (FGBDT: +4.88%, MPRBC: +4.77%, RSLRS: +26.84%).

Keywords: Rough set theory, Decision trees, Related family, Feature selection.

1 INTRODUCTION

Decision trees, first proposed by Quinlan et al. in 1986 [11], are non-cyclic directed graph models based on tree structures, composed of internal nodes and leaf nodes. Internal nodes represent test conditions that can divide data into multiple child nodes based on splitting attributes, while leaf nodes represent decision results [8]. To generate different decision tree structures, researchers have proposed various splitting criteria.

Information Gain (IG) is a splitting criterion used in the ID3 algorithm [11]. However, ID3 has limitations [7]: it only processes categorical features and tends to select features with more values. The C4.5 algorithm [12] introduced Gain Ratio (GR) to overcome these shortcomings, normalizing information gain to avoid bias towards features with more values. The Classification and Regression Tree (CART) algorithm [1]

uses the Gini Index (GI) as a splitting criterion and can handle both classification and regression tasks.

Traditional decision tree algorithms typically use univariate splitting criteria, relying on splitting a single attribute at each node. However, univariate splitting may show limitations with complex decision boundaries or significant attribute interactions. Multi-feature splitting strategies have been proposed to enhance model expressive power [2]. These strategies can represent more complex decision boundaries, require fewer splits resulting in smaller trees, and are particularly effective for high-dimensional data with strong feature interactions [6].

Rough set theory, introduced by Pawlak in 1982 [9], has emerged as a powerful mathematical framework for dealing with vagueness and uncertainty in data. It characterizes concepts through lower and upper approximations, defining a boundary region of uncertainty that makes it particularly suitable for classification problems with imprecise information.

Rough set theory offers distinct advantages for feature selection by working directly with data without requiring additional information like probability distributions. Through analyzing attribute dependencies, it identifies minimal subsets of features (reducts) that preserve classification power. Rough set theory has proven effective for monotonic classification [3], [4], [10], [13] and ensemble learning [5].

Traditionally, Rough set-based feature selection methods are used in data preprocessing before classification. This two-stage "feature selection-classifier application" approach is widely adopted but rarely applied directly to decision tree construction. Recently, Rough set theory has been integrated into classifier development, as seen in Xia et al.'s RSLRS algorithm [14], which uses stability of local redundant attributes for feature selection. However, this method faces challenges with high- dimensional data and noise sensitivity.

Yang et al. proposed the Related Family Method [15], which provides an innovative and efficient solution for feature evaluation in high-dimensional data. This algorithm can remove redundant features while maintaining key discriminative information in the data, reducing computational complexity from exponential or quadratic levels in traditional methods to linear levels. Experiments demonstrate that the Related Family Method significantly outperforms traditional rough set methods across multiple UCI datasets, not only achieving higher classification accuracy (average improvement of 5-10 percentage points) but also using fewer features in some cases. It is precisely because of the out-standing efficiency and effectiveness of the Related Family Method in processing high-dimensional data that it serves as an ideal feature evaluation tool for our proposed Dynamic Related Family Decision Tree (DRFDT) algorithm, providing computational feasibility for dynamically executing feature selection at each node of the decision tree, thus enabling the model to adaptively select optimal splitting feature subsets based on the characteristics of different sample subgroups.

Based on the highly efficient feature evaluation method, we propose the Dynamic Related Family Decision Tree (DRFDT) algorithm, which combines Related Family feature selection efficiency with decision tree interpretability. DRFDT dynamically executes feature selection at each decision node, allowing the model to adaptively select

optimal feature subsets based on local sample characteristics. Our main innovations include:

1. Rough Set-based Dynamic Feature Selection: By integrating the Related Family Method with rough set principles, we propose an adaptive feature selection mechanism that optimizes feature evaluation at each tree node. By constructing maximal consistent granules based on rough set approximations, the algorithm provides node-specific feature selection with strong theoretical guarantees.
2. Adaptive Decision Tree Splitting Mechanism: At each node, the algorithm applies dynamic feature selection based on local sample distribution, enabling better adaptation to varying data patterns. We introduce the "Partition Purity" metric to evaluate candidate feature subsets and apply rough set boundary region analysis to identify and remove noise samples during training.
3. Theoretically Sound Classification Framework: By embedding rough set-based feature selection into the decision tree structure, we maintain interpretability while efficiently processing large- scale datasets. This design avoids the computational burden of traditional decision trees on high-dimensional data while preserving their interpretable advantages.

Through experimental validation, DRFDT outperforms a wide range of classification algorithms across 15 UCI datasets, achieving an average accuracy of 89.30%. This represents significant improvements over classical single-feature decision tree methods (CART: +3.87%), traditional classification algorithms (KNN: +5.71%, SVM: +4.54%), multi-feature split decision tree algorithms (CART-LC: +3.99%, O1: +4.25%), and state-of-the-art decision tree classification algorithms (FGBDT: +4.88%, MPRBC: +4.77%, RSLRS: +26.84%).

2 Rough Set Theory and Dynamic Feature Evaluation Mechanism

2.1 Dynamic Related Family Feature Evaluation

In our previous work [16], we introduced several fundamental concepts related to fuzzy covering decision information systems, including fuzzy covering information systems, lower and upper ap- proximations in fuzzy information systems, positive regions of fuzzy information systems, consistent granules in fuzzy information systems, and attribute reduction methods for fuzzy information systems. Building upon the foundation established in [16], this paper proposes corresponding definitions in the context of crisp sets, and combines them with Related Family [15] method to implement a Dynamic Related feature evaluation approach.

Definition 2.1: Given a nonempty domain U of discourse and the power set $\mathcal{P}(U)$ of U. $\mathcal{C} = \{C_1, C_2, \dots, C_m\}$, where $C_i \in \mathcal{P}(U)$ $(i = 1,2, \dots, m)$ and U is a universal set. $\mathcal{C}$ is called a covering on U, if for every $x \in U$, there exists $C_i \in \mathcal{C}$ such that $x \in C_i$, i.e., $\bigcup_{i=1}^{m} C_i = U$. $(U, \mathcal{C}, D)$ is a covering information system (covering decision system).

Given a dataset data = (U, A, D) whose values are normalized into $[0,1]$ column-by-column, where $U = \{x_1, x_2, \dots, x_n\}$ is the sample set, $A = \{a_1, a_2, \dots, a_m\}$ is the feature set, and $D = \{d_1, d_2, \dots, d_l\}$ is the label set. For $x_i \in U$ and $a_j \in A$, $a_j(x_i)$ represents the value of sample x_i under feature a_j.

1. For convenience, we choose an arithmetic progression as the center point set $G = \{G_k | k = 1,2, \dots, t, G_k = k * \text{step}, t * \text{step} \leq 1\}$, where step is the tolerance.

2. Choose a radius $\delta_{jk} \in [0.1,1]$ to form the granule $C_j^U(G_k, \delta_{jk})$ for each center point $G_k \in G$. Here, j indicates the coverage under feature a_j, and U indicates that the granule is generated for the universe U. The granule is defined as follows:

$$C_j^U(G_k, \delta_{jk}) = \{x_i \in U \mid |a_j(x_i) - G_k| \leq \delta_{jk}\} \qquad (1)$$

$U/D = \{D_1, D_2, \dots, D_h\}$ is the partition of U induced by the equivalence relation D, where each D_i represents an equivalence class of samples with the same decision value.

Definition 2.2: A Covering Information System (CIS) is defined as $(U, \mathcal{C}, D)$, where $\mathcal{C} = \{C_i \mid i = 1,2, \dots, m\}$ is a family of coverings of U, C_i are condition attributes (or the features), and D is the decision attribute (or the label). Let $U/D = \{D_i \mid i = 1,2,3, \dots, l\}$ be the decision classes induced by D, where D is a classical equivalence relation.

Since $\cup \mathcal{C}$ is still a covering on U, for a set $X \subseteq U$, the covering lower approximation on $(U, \mathcal{C}, D)$ can be defined as

$$L_{\cup\mathcal{C}}(X) = \cup \{K \in \cup \mathcal{C} \mid K \subseteq X\}. \qquad (2)$$

The upper approximation can be defined as

$$UP_{\cup\mathcal{C}}(X) = \cup \{K \in \cup \mathcal{C} \mid K \cap X \neq \emptyset\}. \qquad (3)$$

Based on the set inclusion operation, the positive region can be defined as

$$POS_{\mathcal{C}}(D) = \cup \{L_{\cup\mathcal{C}}(D_i) \mid D_i \in U/D\}. \qquad (4)$$

The target of the attribute reduction of a covering information system is to find the minimal subset of $\mathcal{C}$ such that the positive region is invariant.

Definition 2.3: A CIS is defined as $(U, \mathcal{C}, D)$, where $\mathcal{C} = \{C_i \mid i = 1,2, \dots, m\}$ is a family of coverings of U, and $U/D = \{D_i \mid i = 1,2,3, \dots, l\}$. The consistent granule set of $(U, \mathcal{C}, D)$ is defined as

$$Con(U, \mathcal{C}, D) = \{K \mid K \in \cup \mathcal{C}, \exists D_i \in U/D \text{ s.t. } C \subseteq D_i\}. \qquad (5)$$

We term the elements in $Con(U, \mathcal{C}, D)$ as consistent granules of $(U, \mathcal{C}, D)$.

For a center point $G_k \in G$ and the set of all possible radius values $\gamma = \{\delta_1, \delta_2, \dots, \delta_t\}$ where $\delta_i \in [0,1]$, $i = 1,2, \dots, t$, we define the largest consistent granule of G_k as follows: For any $C_h \in \mathcal{C}$, if δ_{jk} is the maximum radius value such that $C_h^U(G_k, \delta_{jk}) \subseteq D_i$ for some $D_i \in U/D$, and there does not exist a value larger than δ_{jk} for which this inclusion holds, then $C_h^U(G_k, \delta_{jk})$ is called the largest consistent granule of G_k relative to the universe U. The collection of all the largest consistent granules under covering C_h is denoted as $\text{MCon}(U, C_h, D)$.

$$\text{MCon}(U, \mathcal{C}, D) = \bigcup_{C_h \in \mathcal{C}} \text{MCon}(U, C_h, D). \qquad (6)$$

Proposition 2.4: $POS_{\mathcal{C}}(D) = \cup \text{M}Con(U, \mathcal{C}, D)$

Definition 2.5: We define reducible features. $(U, \mathcal{C}, D)$ is a CIS, where $\mathcal{C} = \{C_i \mid i = 1,2, \dots, m\}$ is a family of coverings of U, $U/D = \{D_i \mid i = 1,2,3, \dots, l\}$. If $POS_{\mathcal{C}} = POS_{\mathcal{C}-\{C_i\}}$, we say C_i is dispensable in $\mathcal{C}$. Otherwise, C_i is indispensable in $\mathcal{C}$. For every

$\mathbb{P} \subseteq \mathcal{C}$, if $POS_{\mathcal{C}} = POS_{\mathbb{P}}$ and every covering in $\mathbb{P}$ is indispensable, we say $\mathbb{P}$ is a reduct of $\mathcal{C}$.

The collection of all indispensable coverings in $\mathcal{C}$ is denoted by CORE($\mathcal{C}$). The collection of all reducts of $\mathcal{C}$ is denoted by RED($\mathcal{C}$).

Definition 2.6: Dynamic Related Family Feature Evaluation algorithm builds upon the Related Family Feature Evaluation [15]. Given a CIS $(U, \mathcal{C}, D)$, where $\mathcal{C} = \{C_i \mid i = 1,2,\dots,m\}$ is a family of coverings of U, and $MCon(U, \mathcal{C}, D) = \{K_1, K_2, K_3, \dots, K_p\}$, For any $x_i \in U$, we define $r_i = \{C_k \mid x_i \in K_j, K_j \in C_k, C_k \in \mathcal{C}\}$ as the related set of K_j, and $R(U, \mathcal{C}, D) = \{r_i \mid i = 1,2,\dots,m\}$ is the related family of $(U, \mathcal{C}, D)$.

We introduce how to dynamic obtain the related family evaluation results for sample subsets after partitioning the sample set. First, we discuss the monotonicity of the largest consistent granule radius after the sample set is partitioned. Based on this property, we can quickly obtain the largest consistent granules of the sample subsets. Once the largest consistent granules of the sample subsets are obtained, the related family method can be applied to fast determine the feature selection results for the sample subsets.

Property 2.8: Monotonicity Property of Consistent Granule Radius. For any $G_k \in G$, let $C_h^U(G_k, \delta_{ik})$ be the largest consistent granule with respect to the universe U, and $C_h^E(G_k, \delta_{jk})$ be the maximal consistent granule with respect to the universe E. If $E \subseteq U$, then $\delta_{ik} \leq \delta_{jk}$.

Proof. We prove this property by contradiction.

Let us assume the contrary: there exists $E \subseteq U$, but $\delta_{ik} > \delta_{jk}$.

According to the definition of maximal consistent granules, $C_h^U(G_k, \delta_{ik})$ is the largest consistent granule over domain U, and $C_h^E(G_k, \delta_{jk})$ is the largest consistent granule over domain E.

Since $C_h^U(G_k, \delta_{ik})$ is consistent, there exists a decision class D_o such that $C_h^U(G_k, \delta_{ik}) \subseteq D_o$. Given that $E \subseteq U$, for any $x \in E$, if $x \in C_h^U(G_k, \delta_{ik})$, then x must also belong to D_o. This implies that using radius δ_{ik} can also form a consistent granule over domain E: $C_h^E(G_k, \delta_{ik}) \subseteq (D_o \cap E)$.

However, according to our assumption, $\delta_{ik} > \delta_{jk}$, while $C_h^E(G_k, \delta_{jk})$ is the largest consistent granule over E. This creates a contradiction, as we've shown that a consistent granule with a larger radius δ_{ik} could exist over domain E.

Therefore, our initial assumption must be false, and we conclude that $\delta_{ik} \leq \delta_{jk}$. This demonstrates that when we reduce the sample space from U to E for the same center point G_k, the maximum allowable radius for forming consistent granules in the smaller sample space must be greater than or equal to the maximum radius in the original space, as the reduction in sample size decreases the constraints for forming consistent granules.

The monotonicity of consistent granule radius has important applications in decision tree construction, providing both theoretical guarantees and computational advantages. Based on monotonicity, we can design an incremental computation strategy. When moving from parent node u to child node v, for each center point G_k, we only need to start from the parent node's radius δ_u, gradually increase the radius and check consistency, until finding the maximal consistent granule radius δ_v of the child node. This

approach aligns with the incremental learning principle in rough set theory, where knowledge refinement occurs as new information becomes available.

We proposed the maximal consistent granule generation algorithm, applying the monotonicity of maximal consistent granules, to efficiently compute maximal consistent granules.

Table 1. Incremental Maximal Consistent Granule

Algorithm 1 Incremental Maximal Consistent Granule

Require: Maximal consistent granule set $\mathrm{MCon}(U_u, \mathcal{C}, D)$ of parent node u, Sample set U_v of child node v

Ensure: Maximal consistent granule set $\mathrm{MCon}(U_v, \mathcal{C}, D)$ of child node v

Initialize $\mathrm{MCon}(U_v, \mathcal{C}, D) = \emptyset$
for all $N(G_k, \delta_k, D) \in \mathrm{MCon}(S_u, \mathcal{C}, D)$ **do**
    Remove samples not in S_v: $N'(G_k, \delta_k, D) = \{x \in N(G_k, \delta_k, D) | x \in S_v\}$
    Set $\delta = \delta_k$
    while $N'(G_k, \delta, D)$ is consistent and $\delta \leq 1$ **do**
        if $N'(G_k, \delta + 0.01, D)$ is consistent **then**
            $\delta = \delta + 0.01$
        else
            break
        end if
    end while
    Add $N'(G_k, \delta, D)$ to $\mathrm{MCon}(U_v, \mathcal{C}, D)$
end for
Return $\mathrm{MCon}(U_v, \mathcal{C}, D)$

This incremental computation strategy significantly improves the efficiency of feature selection during decision tree construction, enabling our Rough Set-based Dynamic Related Family Decision Trees to maintain low computational complexity on high-dimensional data. By leveraging rough set theory principles, we avoid redundant computations while maintaining the theoretical guarantees on feature selection quality.

In our approach, radius threshold δ serves as a key parameter for controlling the quality of rough set approximations:

Definition 2.9: Adaptive Precision Consistent Granule Generation. Given a covering information system $(U, \mathcal{C}, D)$, where $\mathcal{C} = \{C_i \mid i = 1,2, \dots, m\}$ is a family of coverings of U, and $U/D = \{D_i \mid i = 1,2,3, \dots, l\}$, for each $G_k \in G$, the largest consistent granule set with radius greater than a given threshold β is defined as:

$$\mathrm{MCon}_\beta(U, C_h, D) = \{C_h^U(G_k, \delta_{jk}) \mid C_h^U(G_k, \delta_{jk}) \in \mathrm{MCon}(U, C_h, D), G_k \in G, \delta_{jk} > \beta\}. \quad (7)$$

Thus, the largest consistent granule set with radius greater than β for the entire covering system $\mathcal{C}$ is:

$$\mathrm{MCon}_\beta(U, \mathcal{C}, D) = \bigcup_{C_h \in \mathcal{C}} \mathrm{MCon}_\beta(U, C_h, D). \quad (8)$$

After obtaining the maximal consistent granule set of the node through threshold parameters, we can derive feature evaluation results for the node using the related family method. The specific algorithmic process is as follows:

Table 2. Dynamic Related Family Feature Evaluation

Algorithm 2 Dynamic Related Family Feature Evaluation

Require: Covering information system $(U, \mathcal{C}, D)$, radius threshold β
Ensure: Feature subset Δ (approximate reduct)
Initialize $\Delta = \emptyset$
Obtain all largest consistent granules $\mathrm{MCon}(U, \mathcal{C}, D)$
Filter granules by radius threshold: $\mathrm{MCon}_\beta(U, \mathcal{C}, D) = \{N \in \mathrm{MCon}(U, \mathcal{C}, D) | \mathrm{radius}(N) > \beta\}$
Compute positive region $\mathrm{POS}_\mathcal{C}$ by taking the union of $\mathrm{MCon}_\beta(U, \mathcal{C}, D)$
for all $x_i \in \mathrm{POS}_\mathcal{C}$ **do**
    Construct its related set $r_i = \{K_j | x_i \in K_j, K_j \in C_k, C_k \subseteq \mathcal{C}\}$
end for
Form the related family $R(U, \mathcal{C}, D) = \{r_i | x_i \in \mathrm{POS}_\mathcal{C}\}$
while $R(U, \mathcal{C}, D) \neq \emptyset$ **do**
    Count frequency of features in $R(U, \mathcal{C}, D)$
    Find feature C_k with highest frequency
    $\Delta = \Delta \cup \{C_k\}$
    Remove all r_i containing C_k from $R(U, \mathcal{C}, D)$
end while
Return Δ

This rough set reformulation not only provides theoretical justification for the Related Family method but also establishes it as a principled approach for feature selection within the well-established framework of rough set theory

3 Multi-Feature Splitting

In our Rough Set-based Dynamic Related Family Decision Trees, samples are projected onto lower-dimensional spaces through feature subsets selected based on rough set principles. This projection process can be formalized as follows:

Definition 3.1: Feature Subset Projection. Let F'_i be a feature subset obtained by Dynamic Related Family Feature Evaluation algorithm. The projection of the dataset U onto F'_i is defined as:

$$U'_i = U \cdot M_{F'_i} \qquad (9)$$

where $M_{F'_i}$ is a projection matrix that retains only the columns corresponding to the features in F'_i. The resulting dataset U'_i has N samples and $|F'_i|$ features.

From a rough set perspective, this projection operation preserves the discernibility relations determined to be most important by the feature selection process. If the feature subset F'_i is a reduct or an approximate reduct of the original feature set, then the projection preserves the positive region and maintains essential classification information.

Partition Purity serves as a key metric in our approach, quantifying how well the feature subset approximates the decision classes. From a rough set perspective, this metric evaluates the quality of the lower and upper approximations:

Definition 3.2: Partition Purity for Feature Subsets. Let $\{C_1, C_2, \ldots, C_q\}$ be the clusters obtained from K-means clustering on U'_i, and let $\{L_1, L_2, \ldots, L_w\}$ be the true label sets. The Partition Purity of the feature subset F'_i is defined as:

$$\text{Partition Purity}(F'_i) = \frac{1}{N} \sum_{j=1}^{k} \max_{\ell} \left| C_j \cap L_\ell \right| \qquad (10)$$

where $|C_j \cap L_\ell|$ is the number of samples in cluster C_j that have the true label L_ℓ, and $\max_\ell |C_j \cap L_\ell|$ represents the majority label count in cluster C_j.

The value range of Partition Purity is $[0,1]$, where 1 represents perfect clustering (each cluster contains samples of only one class, equivalent to perfect lower approximation), and lower values indicate increasing approximation error.

This rough set interpretation of Partition Purity provides a theoretical foundation for evaluating feature subsets based on their approximation quality, guiding the selection of optimal splitting features at each node.

Our approach innovatively integrates K-means clustering with rough set principles to create a multi-path splitting mechanism. These theoretical advantages derive from the synergy between rough set theory's focus on discernibility relations and clustering's ability to identify natural groupings in the feature space. In our Dynamic Related Family Decision Tree approach, the selection of the number of clusters K is straightforward and practical–we simply set K equal to the number of classes in the dataset, where $|Y|$ represents the number of distinct classes in the dataset. Our experiments across diverse datasets confirm that this straightforward approach works effectively in practice. Even for datasets with a large number of classes (such as the Movement libras dataset with 15 classes), setting K equal to the class count produces well-balanced splits that lead to high classification accuracy. The simple relationship between classes and clusters aligns well with the fundamental classification objective–separating samples from different classes into distinct groups. By matching the cluster count to the class count, we create a natural structural correspondence between the problem definition and the algorithmic approach.

4 Construction and Prediction of Rough Set-Based Dynamic Related Family Decision Trees

4.1 Rough Set-Based Dynamic Related Family Decision Tree Construction Algorithm

The complete algorithm for constructing our Rough Set-based Dynamic Related Family Decision Tree integrates rough set principles with multi-feature splitting:

Table 3. Build Dynamic Related Family Decision Tree

Algorithm 3 Build Dynamic Related Family Decision Tree

Require: Training dataset (U, A, D), minimum sample threshold T_1, purity threshold T_2, radius threshold β
Ensure: Root node of the Rough Set-based Dynamic Related Family Decision Tree

function BUILDTREE(S, depth)
  $v \leftarrow$ new Node
  $P(v) \leftarrow \max_{c \in Y} \frac{|\{x \in S | \text{label}(x) = c\}|}{|S|}$ ▷ Calculate node purity
  if $|S| < T_1$ **or** $P(v) > T_2$ **or** $P(v) = 1$ **or** depth > maxDepth **then**
    v.prediction $\leftarrow \arg\max_{c \in Y} |\{x \in S \mid \text{label}(x) = c\}|$
    v.isLeaf $\leftarrow$ true
    return v
  end if
  $(U, \mathcal{C}, D) \leftarrow$ transform dataset S to covering information system
  MCon$(U, \mathcal{C}, D) \leftarrow$ compute maximal consistent granules
  candidateSubsets $\leftarrow \emptyset$ ▷ Generate candidate feature subsets
  for all $\delta_i \in \{\delta - 0.1, \delta - 0.05, \delta, \delta + 0.05, \delta + 0.1\}$ **do**
    if $\delta_i > 0$ **and** $\delta_i \leq 1$ **then**
      $F_i' \leftarrow$ DynamicRelatedFamilyFeatureEvaluation$(U, \mathcal{C}, D, \delta_i)$ ▷ Algorithm 2
      if F_i' is not empty **then**
        candidateSubsets $\leftarrow$ candidateSubsets $\cup \{F_i'\}$
      end if
    end if
  end for
  if candidateSubsets is empty **then**
    v.prediction $\leftarrow \arg\max_{c \in Y} |\{x \in S \mid \text{label}(x) = c\}|$
    v.isLeaf $\leftarrow$ true
    return v
  end if
  bestSubset $\leftarrow$ null, bestPurityGain $\leftarrow 0$, bestClusters $\leftarrow$ null, bestCenters $\leftarrow$ null
  for all $F_i' \in$ candidateSubsets **do**
    $S'_{\text{proj}} \leftarrow$ project S' onto feature subset F_i'
    $k \leftarrow |Y|$, $\{C_1, C_2, \ldots, C_k\}, \{c_1, c_2, \ldots, c_k\} \leftarrow$ K-means(S'_{proj}, k)
    for all $j = 1$ to k **do**
      $P(C_j) \leftarrow \max_{c \in Y} \frac{|\{x \in C_j | \text{label}(x) = c\}|}{|C_j|}$
    end for
    PurityGain$_i \leftarrow \sum_{j=1}^{k} \frac{|C_j|}{|S'|} \cdot (P(C_j) - P(v))$
    if PurityGain$_i >$ bestPurityGain **then**
      bestPurityGain $\leftarrow$ PurityGain$_i$
      bestSubset $\leftarrow F_i'$
      bestClusters $\leftarrow \{C_1, C_2, \ldots, C_k\}$
      bestCenters $\leftarrow \{c_1, c_2, \ldots, c_k\}$
    end if
  end for
  if bestPurityGain < minGain **then**
    v.prediction $\leftarrow \arg\max_{c \in Y} |\{x \in S \mid \text{label}(x) = c\}|$
    v.isLeaf $\leftarrow$ true
    return v
  end if
  v.featureSubset $\leftarrow$ bestSubset
  v.clusterCenters $\leftarrow$ bestCenters
  for all $i = 1$ to k **do**
    if $|\text{bestClusters}[i]| > 0$ **then**
      MCon(bestClusters$[i], \Gamma, D) \leftarrow$ IncrementalMaximalConsistentGranule(MCon(U, Γ, D), bestClusters$[i]$) ▷ Algorithm 1
      v.children$[i] \leftarrow$ BuildTree(bestClusters$[i]$, depth $+ 1$)
    else
      v.children$[i] \leftarrow$ NULL
    end if
  end for
  return v
end function
return BUILDTREE(U, 0)

4.2 Rough Set-Based Prediction Algorithm

The prediction algorithm implements the navigation through approximation spaces described in the previous section:

Table 4. Predict With Dynamic Related Family Decision Tree

Algorithm 4 Predict With Dynamic Related Family Decision Tree

Require: Root node of a trained Rough Set-based Dynamic Related Family Decision Tree, test sample x

Ensure: Predicted class label for x

Function Predict(node, x):
if node.isLeaf **then**
    Return node.prediction
end if
    // Project test sample onto node's approximation space
    $x_{proj} \leftarrow x \cdot M_{\text{node.featureSubset}}$
    // Find the nearest approximation region center using discernibility measure
    $\text{next}(v, x) = \arg\min_i d(x|_{F'_v}, c_i)$
    where $d(x, c_i) = ||x|_{F'v} - c_i||_2 = \sqrt{\sum_{f \in F'_v} (f(x) - f(c_i))^2}$
    nextNode $\leftarrow$ node.children[next(v, x)]
if nextNode is NULL **then**
    Return node.prediction // Fallback to current approximation
end if
    Return Predict(nextNode, x)
Return Predict(root, x)

The prediction process navigates through a hierarchy of approximation spaces, identifying the most appropriate refinement at each step based on the discernibility between the test sample and approximation centers. This rough set-based prediction approach enables efficient classification of new samples while maintaining the theoretical guarantees provided by the rough set framework

5 EXPERIMENTAL ANALYSIS

In this section, we conduct comprehensive experiments to evaluate the classification performance of our proposed Rough Set-based Dynamic Related Family Decision Tree (DRFDT) algorithm. Our experimental analysis aims to validate the practical effectiveness of the rough set theoretical framework we've established and demonstrate its advantages in real-world classification tasks. We compare DRFDT against three classical classification algorithms and three state-of-the-art decision tree variants. Additionally, we investigate the impact of our adaptive parameter selection mechanism on classification performance. We carefully selected fifteen datasets from the UCI machine learning repository, covering a diverse range of domains, dimensions, and complexities, as detailed in Table 5. Prior to experimentation, all datasets underwent preprocessing to normalize numerical features to the [0,1] range, facilitating consistent granule construction and rough set approximation.

Table 5. Description of experimental datasets.

No	Data Sets	Sample	Features	Classes
1	aliz v3	62	2094	4
2	leukemia	72	7071	2
3	armstrong	72	2195	3
4	CLL_SUB	111	11341	3
5	inosphere	351	35	2
6	Movement libras	360	91	15
7	Ultrasonic flowmeter diagnostics	361	44	4
8	DrivFace	606	6401	3
9	vowel	990	14	11
10	Obesity Estimation	2111	17	7
11	segment	2310	20	7
12	twonorm	7400	21	2
13	thyroid	7200	22	3
14	electrical	10000	14	2
15	Internet Firewall Data	65532	12	4

Our experimental evaluation compares DRFDT with three classical classification algorithms—Classification and Regression Trees (CART), Support Vector Machines (SVM), and k-Nearest Neighbors (KNN with k=3)—as well as three state-of-the-art decision tree variants: Feature Generalized Binary Decision Tree (FGBDT), Multi-Perspective Rule-Based Classification (MPRBC), and Relative Stability Local Redundancy-based Rough Set (RSLRS). We also included two additional tree-based methods: Oblique Decision Tree (O1) and CART with Linear Combinations (CART-LC), which employ multivariate splitting strategies. To comprehensively evaluate classification performance, we adopted accuracy as the primary metric and employed ten-fold cross-validation across all experiments.

In Our experimental, A key mechanism in our DRFDT algorithm is its built-in adaptive mechanism. This mechanism divides the training sample set into an internal training set (90%) and an internal validation set (10%). For each candidate leaf node purity threshold, ranging from 0.90 to 0.99 with increments of 0.01, the algorithm evaluates classification accuracy. The leaf node purity threshold that yields the highest classification accuracy is selected as the optimal value for the dataset. The DRFDT tree is then trained on the entire training sample set using this optimized threshold.

5.1 Performance Comparison and Analysis

Table 6 presents the classification accuracy of all algorithms on the 15 datasets. For fair comparison, we employed consistent 10-fold cross-validation and report the average accuracy percentages. The highest accuracy for each dataset is highlighted in red (with values within 1% of the maximum considered equivalent), while the second-highest is in blue.

Table 6. Classification Accuracy Comparison on UCI Datasets (%).

Dataset	SVM	CART	KNN	FGBDT	MPRBC	RSLRS	O1	CART-LC	DRFDT
vowel	86.87	78.08	60.71	64.34	61.71	55.05	55.56	58.08	86.36
Ultrasonic	50.13	84.77	73.95	83.10	74.95	42.11	65.75	82.19	89.19
twonorm	97.70	84.81	96.62	81.31	97.89	70.45	79.73	93.31	97.66
thyroid	93.13	99.69	93.81	99.49	94.81	71.14	99.38	99.24	96.79
segment	93.07	96.67	94.94	90.91	95.94	68.01	93.51	83.77	95.54
Obesity	92.80	97.30	97.30	91.71	93.33	56.80	88.89	87.71	96.54
libras	79.72	61.39	68.33	58.33	69.33	58.06	40.28	44.44	76.39
leukemia	84.48	94.38	84.76	94.46	85.76	72.22	72.23	70.18	93.05
Firewall	92.02	99.77	99.39	99.85	99.39	54.93	99.85	99.82	99.30
inosphere	93.44	88.04	83.49	89.73	84.49	72.08	85.92	85.92	88.93
electrical	98.15	99.98	91.86	99.99	92.86	77.81	100	100	99.11
DrivFace	90.10	94.56	95.87	100	96.87	74.75	100	100	95.55
CLL_SUB	35.14	62.17	54.94	66.67	55.94	51.35	99.38	98.94	56.72
armstrong	95.89	75.18	82.14	75.18	83.14	75.00	97.04	97.22	86.43
aliz v3	88.81	63.81	80.48	71.19	81.48	37.10	98.33	97.85	81.90
Average	84.76	85.44	83.59	84.42	84.53	62.46	85.05	85.31	89.30

The experimental results demonstrate that our proposed DRFDT algorithm, grounded in rough set theory, achieves superior overall performance across the tested datasets. DRFDT attains the highest average classification accuracy (89.30%) among all compared methods, significantly outperforming both classical algorithms (SVM: 84.76%, CART: 85.44%, KNN: 83.59%) and state-of-the-art decision tree variants (FGBDT: 84.42%, MPRBC: 84.53%, RSLRS: 62.46%, O1: 85.05%, CART-LC: 85.31%). DRFDT outperforms a wide range of classification algorithms across 15 UCI datasets, achieving an average accuracy of 89.30%, which represents an average improvement of 7.35% compared to other methods. This represents significant improvements over classical single-feature decision tree methods (CART: +3.87%), traditional classification algorithms (KNN: +5.71%, SVM: +4.54%), multi-feature split decision tree algorithms (CART-LC: +3.99%, O1: +4.25%), and state-of-the-art decision tree classification algorithms (FGBDT: +4.88%, MPRBC: +4.77%, RSLRS: +26.84%). DRFDT provides more balanced and consistent performance across a wider range of datasets. This suggests that our rough set-based approach to feature selection and decision boundary construction offers a more generalizable classification framework, particularly effective when handling datasets with moderate dimensionality and complex class distributions.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we present the Rough Set-based Dynamic Related Family Decision Tree (DRFDT), a novel decision tree algorithm that integrates rough set theory with adaptive feature selection and dynamic sample partitioning. Our experimental results across 15 diverse UCI datasets demonstrate that DRFDT significantly outperforms both classical machine learning algorithms and state-of-the-art decision tree methods, achieving an average classification accuracy of 89.30%. The statistical analysis confirms that DRFDT provides substantial improvements in classification performance while maintaining the interpretability advantages of decision tree models.

For future work, we plan to explore several promising directions to further enhance the DRFDT algorithm:

1. **Rough Set-Based Adaptive Clustering**: While our current implementation utilizes k-means for sample partitioning, future research will focus on developing a clustering approach directly based on rough set approximation regions. This would leverage the natural granularity of rough set theory to determine optimal partitioning without requiring predefined cluster numbers, potentially providing more semantically meaningful decision boundaries.

2. **Variable Precision Rough Set Integration**: We aim to formally incorporate variable precision rough set models that allow controlled relaxation of approximation precision. This would provide a more flexible theoretical framework for handling datasets with different noise levels and class boundary characteristics.

The rough set perspective offers a theoretically sound foundation for decision tree construction that differs fundamentally from traditional impurity-based splitting criteria. By establishing a formal connection between rough set approximation quality and decision tree optimization, our work opens new avenues for research at the intersection of these fields. We believe that further development of these theoretical connections will lead to significant advances in both the efficiency and effectiveness of decision tree models for complex classification tasks.

References

1. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. CRC Press (1984)
2. Brodley, C.E., Utgoff, P.E.: Multivariate Decision Trees. Machine Learning 19, 45-77 (1995)
3. Hu, Q., Che, X., Zhang, L., Zhang, D., Guo, M., Yu, D.: Rank Entropy-Based Decision Trees for Monotonic Classification. IEEE Transactions on Knowledge and Data Engineering 24(11), 2052-2064 (2011)
4. Hu, Q., Pan, W., Zhang, L., Zhang, D., Song, Y., Guo, M., Yu, D.: Feature Selection for Monotonic Classification. IEEE Transactions on Fuzzy Systems 20(1), 69-81 (2011)
5. Hu, Q., Yu, D., Xie, Z., Li, X.: EROS: Ensemble Rough Subspaces. Pattern Recognition 40(12), 3728-3739 (2007)

6. Lim, T.S., Loh, W.Y., Shih, Y.S.: A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. Machine Learning 40(3), 203-228 (2000)
7. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann, San Francisco (2011)
8. Mu, Y., Liu, X., Wang, L.: A Pearson's Correlation Coefficient Based Decision Tree and Its Parallel Implementation. Information Sciences 435, 40-58 (2018)
9. Pawlak, Z.: Rough Sets. International Journal of Computer & Information Sciences 11(5), 341-356 (1982)
10. Qian, Y., Xu, H., Liang, J., Liu, B., Wang, J.: Fusing Monotonic Decision Trees. IEEE Transactions on Knowledge and Data Engineering 27(10), 2717-2728 (2015)
11. Quinlan, J.R.: Induction of Decision Trees. Machine Learning 1(1), 81-106 (1986)
12. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
13. Wang, J., Qian, Y., Li, F., Liang, J., Ding, W.: Fusing Fuzzy Monotonic Decision Trees. IEEE Transactions on Fuzzy Systems 28(5), 887-900 (2019)
14. Xia, S., Bai, X., Wang, G., Cheng, Y., Meng, D., Gao, X., Zhai, Y., Giem, E.: An Efficient and Accurate Rough Set for Feature Selection, Classification, and Knowledge Representation. IEEE Transactions on Knowledge and Data Engineering 35(8), 7724-7735 (2022)
15. Yang, T., Li, Q., Zhou, B.: Related Family: A New Method for Attribute Reduction of Covering Information Systems. Information Sciences 228, 175-191 (2013)
16. Yang, T., Li, Y.J., Qian, Y., Wang, F.Y.: Consistent Matrix: A Feature Selection Framework for Large-Scale Datasets. IEEE Transactions on Fuzzy Systems 31(11), 4024-4038 (2023)