electronics

MDPI

Article

Research on a Decision Tree Classification Algorithm Based on Granular Matrices

Lijuan Meng [1], Bin Bai [1,2,3,4,*], Wenda Zhang [5], Lu Liu [1,2,3,4,6,*] and Chunying Zhang [1,2,3,4,6]

[1] College of Science, North China University of Science and Technology, Tangshan 063210, China; menglijuan@stu.ncst.edu.cn (L.M.); zchunying@ncst.edu.cn (C.Z.)
[2] Hebei Engineering Research Center for the Intelligentization of Iron Ore Optimization and Ironmaking Raw Materials Preparation Processes, North China University of Science and Technology, Tangshan 063210, China
[3] Hebei Key Laboratory of Data Science and Application, North China University of Science and Technology, Tangshan 063210, China
[4] The Key Laboratory of Engineering Computing in Tangshan City, North China University of Science and Technology, Tangshan 063210, China
[5] College of Mining Engineering, North China University of Science and Technology, Tangshan 063210, China; zhangwenda@stu.ncst.edu.cn
[6] Tangshan Intelligent Industry and Image Processing Technology Innovation Center, North China University of Science and Technology, Tangshan 063210, China
* Correspondence: baibin@ncst.edu.cn (B.B.); liulu_hblg@ncst.edu.cn (L.L.)

Abstract: The decision tree is one of the most important and representative classification algorithms in the field of machine learning, and it is an important technique for solving data mining classification tasks. In this paper, a decision tree classification algorithm based on granular matrices is proposed on the basis of granular computing theory. Firstly, the bit-multiplication and bit-sum operations of granular matrices are defined. The logical operations between granules are replaced by simple multiplication and addition operations, which reduces the operation time. Secondly, the similarity between granules is defined, the similarity metric matrix of the granular space is constructed, the classification actions are extracted from the similarity metric matrix, and the classification accuracy is defined by weighting the classification actions with the probability distribution of the granular space. Finally, the classification accuracy of the conditional attribute is used to select the splitting attributes of the decision tree as the nodes to form forks in the tree, and the similarity between granules is used to judge whether the data types in the sub-datasets are consistent to form the leaf nodes. The feasibility of the algorithm is demonstrated by means of case studies. The results of tests conducted on six UCI public datasets show that the algorithm has higher classification accuracy and better classification performance than the ID3 and C4.5.

Keywords: classification; decision tree; granular computing; granular structure; granular matrix; similarity metric matrix; classification accuracy

Citation: Meng, L.; Bai, B.; Zhang, W.; Liu, L.; Zhang, C. Research on a Decision Tree Classification Algorithm Based on Granular Matrices. *Electronics* **2023**, *12*, 4470. https://doi.org/10.3390/electronics12214470

Academic Editor: Andrei Kelarev

Received: 11 September 2023
Revised: 22 October 2023
Accepted: 24 October 2023
Published: 30 October 2023

1. Introduction

Classification is not only a common problem in life but also a hot problem for research in fields such as machine learning and data mining. The main task of classification is to build a model that reflects the mapping relationship between samples and labels by generalizing and learning a series of sample data with labels [1]. The decision tree is one of the most important and representative classification algorithms, and it is an important technique to solve classification tasks in data mining. With the advantages of fast classification speed, a simple rule generation process, and strong interpretive power, the decision tree has received broad interest.

The core idea of the decision tree algorithm is to find the appropriate splitting attributes as the nodes of the tree, divide the dataset to form a fork in the tree, and recursively call the forking process for each subset of data. The final decision tree is generated after all

subsets contain the same type of data. The constructed decision tree can then classify new data samples. The key to the decision tree classification algorithm is selecting the splitting attributes. The information gain, Taylor's formula, the Gini coefficient, and granular computing are the most commonly used attribute selection operators. The splitting attributes were selected through information gain and its improvements in classical ID3, C4.5, E-ID3, R-C4.5, etc. To simplify the logarithmic process, many mathematical methods have been used in the literature to select the splitting attributes, such as Taylor's formula, McLaughlin's formula, and equivalence substitution instead of the entropy function [2,3]. The splitting attributes were selected through the Gini coefficient and its improvements in CART [4], SLIQ [5], SPRINT [6], FS-DT [7], model decision trees [8], etc. The splitting attributes were selected through granular computing in FDT [9], IFDT [10], decision trees based on fuzzy sets [11], ordered decision trees [12,13], and decision tree algorithms based on approximation accuracy [14,15], the purity of attributes [16], and the importance of attributes [17], among others. In order to further optimize the classification performance of decision trees, many optimized decision tree classification models have been proposed through the integration of commonly used attribute selection operators with other theories. Examples include a decision tree integrating information gain and the Gini index [18], a decision tree integrating information entropy and the correlation coefficient [19] or the covariance [20], a decision tree based on ant colony optimization [21], and a decision tree with parallelization [22]. These optimized decision tree models extended the decision tree classification method and improved decision tree theory.

Granular computing is a new computing paradigm in the current field of intelligent information processing [23] and is a new method of solving complex problems with multi-granularity, multi-perspective, and multi-level aspects in the form of knowledge granules. Granular computing has attracted much attention because it is in line with the basic idea of human processing and solving complex problems [24]. Granular computing plays an important role in decision tree classification algorithms such as fuzzy sets, rough sets, word computation, quotient spaces, cloud models, etc. [25–28]. The granular matrix [29] is a new mathematical model proposed on the theoretical basis of granular computing. A granular matrix is a binary representation of the granular structure that presents the classification ability of knowledge and can well solve problems with an incomprehensible nature and poorly intuitive operations.

Based on the above analysis, in this paper, we aim to integrate the algebraic operations of granular matrices into the decision tree algorithm, propose a decision tree classification algorithm based on granular matrices, and give a new decision tree division method based on the similarity measure matrix of the granular structure. The test results show that the algorithm has higher classification accuracy and better classification performance than the ID3 and C4.5.

The main innovative works of this paper are as follows:

(1) A division of attributes is considered as a granular structure. The "bit-multiplication" and "bit-sum" operations of the granular matrix are defined. The similarity measure matrix of the granular structure is given.
(2) The method of extracting the classification decision matrix and calculating the classification accuracy of conditional attributes is given and used as the principle of decision tree splitting. This is a new decision tree construction method.
(3) The process is simplified by the operation mode of binary granules that applies from the mathematical operation of theoretical description to the practical operation of the computer model.

The article is structured as follows.

In Section 2, the granular structure of the decision information system is introduced in detail, these operation rules of the "bit-multiplication" and "bit-sum" are given, and the similarity measure matrix of the granular structure is given. In Section 3, the decision tree classification algorithm based on granular matrices is introduced in detail, and the efficiency of the algorithm is analyzed. In Section 4, a comparative test is given between the

presented algorithm and the classical ID3 and C4.5; the test results are analyzed in detail. Finally, a summary is given for the decision tree classification algorithm based on granular matrices, and subsequent research contents are proposed.

2. Basic Concepts

Granular computing is recognized as a new conceptual and computational paradigm for information processing [30]. Granulation, granules, and granularity are the most fundamental concepts in granular computing. Granulation is the division of a theoretical domain by indistinguishable relations, a granule is a block formed by this division, and granularity is a homogenized measure of the coarseness of information.

2.1. Information System

Definition 1 [31]. *An information system is of the form* $K = (U, AT, V, f)$*, abbreviated as* $K = (U, AT)$*, where* U *is a dataset, referred to as the domain;* AT *is the attribute set; and* $V = \cup_{a \in AT} V_a$ *is the domain of values, i.e.,*$\forall x \in U, a \in AT$*, we have* $f_a(x) \in V_a$*.* $f : U \to V$ *is called the information function.*

In particular, if AT contains the conditional attributes set C and the decision attributes set D, i.e., $AT = C \cup D, C \cap D = \phi$, then $K = (U, AT)$ is a decision information system or decision table, abbreviated as $K = (U, C \cup D)$.

For example, Table 1 presents a decision information system. Here, $U = \{1, 2, 3, 4, 5, 6\}$, conditional attributes set $C = \{color, price, size\}$, and decision attributes set $D = \{Buy\}$.

Table 1. A decision information system, K = (U,C∪D).

U	C			D
	Color	**Price**	**Size**	**Buy**
1	White	High	Full	No
2	White	Low	Compact	Yes
3	Black	Low	Compact	No
4	Black	Low	Full	Yes
5	Grey	High	Full	Yes
6	Grey	High	Compact	No

Definition 2 [31]. *In* $K = (U, AT)$*,* $\forall R \subseteq AT$*, an indistinguishable relation* $IND(R)$ *is defined:* $IND(R) = \{(x, y) \in U \times U | \forall a \in R, f_a(x) = f_a(y)\}$*. The domain* U *is divided into* $U/IND(R) = \{X_1^R, X_2^R, \cdots, X_m^R\}, 1 \leq m \leq |U|$*, under the effect of* $IND(R)$*. Here,* X_i^R *is the information granule determined by R, abbreviated as a granule;* $U/IND(R)$ *is the granular structure;* $P_i^R = |X_i^R|/|U|$ *is the probability distribution of* $U/IND(R)$*; and* $|*|$ *denotes the number of elements contained in the granule.*

For example, if $R = \{Colour\}$, then we have the granular structure $U/IND(R) = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$; $X_R^1 = \{1, 2\}, X_R^2 = \{3, 4\}$, and $X_R^3 = \{5, 6\}$ are the information granules determined by R; and $P_i^R = |X_i^R|/|U| = 1/2, i = 1, 2, 3$, from Table 1.

In particular, in a decision information system $K = (U, C \cup D)$, X_i^P is the conditional granule determined by $P \subseteq C$, and X_i^Q is the decision granule determined by $Q \subseteq D$.

Definition 3 [32]. *In* $K = (U, AT)$*,* $U/IND(R) = \{X_1^R, X_2^R, \cdots, X_m^R\}$ *is the granular structure determined by* $R \subseteq AT$*, and* $GD(R) = \sum_{i=1}^{m} |X_i^R|^2 / |U|^2$ *is called the granularity of R.*

For example, if $R = \{Colour\}$, then $GD(R) = \sum_{i=1}^{m} \left|X_i^R\right|^2 / |U|^2 = \left(2^2 + 2^2 + 2^2\right)/6^2 = 1/3$, from Table 1.

In granular computing, information systems are also called knowledge bases, and attributes are called knowledge considered as a classification capability. The stronger the classification capability of the knowledge (attributes R), the finer the division of the domain, the smaller the granularity $GD(R)$, and the more important the attributes R. Particularly, if $U/IND(R) = \left\{u_1, u_2, \cdots, u_{|U|}\right\}$, then $U/IND(R)$ is the finest division of the domain U, and $GD(R) = 1/|U|$. If $U/IND(R) = U$, then $U/IND(R)$ is the coarsest division of the domain U, and $GD(R) = 1$.

2.2. The Granular Matrix and Its Operations

The granular structure and the classification ability of knowledge are intuitively presented in the form of a Boolean matrix in granular computing, which can well solve these problems of knowledge with incomprehensible nature and poorly intuitive operations. The "bit-multiplication" and "bit-sum" operations of the granular matrix are defined, instead of intersection and concatenation operations on the information granules, in this section.

Definition 4. *In $K = (U, AT)$, $U/IND(R) = \left\{X_1^R, X_2^R, \cdots, X_m^R\right\}, 1 \le m \le |U|$ is the granular structure determined by $R \subseteq AT$. If the mapping $x_{ij}^R = \begin{cases} 1, & u_j \in X_i^R \\ 0, & u_j \notin X_i^R \end{cases}, j = 1, 2, \cdots, |U|$ is defined, then X_i^R can be represented by a Boolean row $\left(x_{i1}^R, x_{i2}^R, \cdots, x_{i|U|}^R\right)$, i.e., $\overrightarrow{X_i^R} = \left(x_{i1}^R, x_{i2}^R, \cdots, x_{i|U|}^R\right)$. $M^R = \left(\overrightarrow{X_1^R}; \overrightarrow{X_2^R}; \cdots; \overrightarrow{X_m^R}\right)$ is then the granular matrix induced by R.*

For example, if $R = \{Colour\}$, then $\overrightarrow{X_1^R} = (1,1,0,0,0,0)$, $\overrightarrow{X_2^R} = (0,0,1,1,0,0)$, $\overrightarrow{X_3^R} = (0,0,0,0,1,1)$, and $M^R = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$, from Table 1.

In particular, if $GD(R) = 1$, then $M^R = (1, 1, \cdots, 1)$ is a row vector; if $GD(R) = 1/|U|$, $M^R = I$ is a unit matrix.

In $K = (U, AT)$, suppose that $M^P = \left(\overrightarrow{X_1^P}; \overrightarrow{X_2^P}; \cdots; \overrightarrow{X_m^P}\right)$ and $M^Q = \left(\overrightarrow{X_1^Q}; \overrightarrow{X_2^Q}; \cdots; \overrightarrow{X_n^Q}\right)$ are granular matrices induced by $P, Q (P, Q \subseteq AT)$, respectively, where $\overrightarrow{X_i^P} = \left(x_{i1}^P, x_{i1}^P, \cdots, x_{i|U|}^P\right)$ and $\overrightarrow{X_j^Q} = \left(x_{j1}^Q, x_{j1}^Q, \cdots, x_{j|U|}^Q\right), 1 \le i \le m \le |U|, 1 \le j \le n \le |U|$.

The "bit-multiplication" and "bit-sum" operations of these granular matrices are defined as follows:

Definition 5. $M^P \otimes M^Q = \left(\overrightarrow{X_i^P} \otimes \overrightarrow{X_j^Q}\right)_{mn \times |U|}$ *is called "bit-multiplication" between the granular matrices M^P and M^Q, where $\overrightarrow{X_i^P} \otimes \overrightarrow{X_j^Q} = \left(x_{i1}^P \cdot x_{j1}^Q, x_{i2}^P \cdot x_{j2}^Q, \cdots, x_{i|U|}^P \cdot x_{j|U|}^Q\right)$.*

In the "bit-multiplication, $\otimes$"operation between M^P and M^Q, all elements of each row in M^P are multiplied by the corresponding elements of all rows in M^Q. Each row vector $\overrightarrow{X_i^P} \otimes \overrightarrow{X_j^Q}$ in $M^P \otimes M^Q$ is identical to the Boolean row vector of the information granules $X_i^P \cap X_j^Q$. The matrix $M^P \otimes M^Q$ obtained through the "bit-multiplication" opera-

tion between M^P and M^Q is identical to the granular matrix $M^{P\cup Q}$ induced by $P \cup Q$, i.e., $M^{P\cup Q} = M^P \otimes M^Q$.

For example, if $P = \{Price\}$ and $Q = \{Size\}$, then $\overrightarrow{X_1^P} = (1,0,0,0,1,1)$, $\overrightarrow{X_2^P} = (0,1,1,1,0,0)$, $\overrightarrow{X_1^Q} = (1,0,0,1,1,0)$, $\overrightarrow{X_2^Q} = (0,1,1,0,0,1)$, $M^P = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$, and $M^Q = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$, from Table 1.

Based on Definition 5, $M^P \otimes M^Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$.

Definition 6. $M^P \oplus M^Q = \left(\overrightarrow{X_i^P} + \overrightarrow{X_j^Q}\right)_{mn\times|U|}$ *is called the "bit-sum" between the granular matrices* M^P *and* M^Q, *where* $\overrightarrow{X_i^P} + \overrightarrow{X_j^Q} = \left(x_{i1}^P + x_{j1}^Q, x_{i2}^P + x_{j2}^Q, \cdots, x_{i|U|}^P + x_{j|U|}^Q\right)$.

In the "bit-multiplication, $\oplus$"operation between M^P and M^Q, all elements of each row in M^P are added to the corresponding elements of all rows in M^Q. The row vector $\overrightarrow{X_i^P} + \overrightarrow{X_j^Q}$ is the element of $M^P \oplus M^Q$ in row $(i-1)n + j$. $\overrightarrow{X_i^P} + \overrightarrow{X_j^Q} - \overrightarrow{X_i^P} \otimes \overrightarrow{X_j^Q}$ is the element of $M^P \oplus M^Q - M^P \otimes M^Q$ in row $(i-1)n + j$ and is consistent with the row vector of the information granules $X_i^P \cup X_j^Q$.

For example, if $P = \{Price\}$ and $Q = \{Size\}$, then based on Definition 6, $M^P \oplus M^Q = \begin{pmatrix} 2 & 0 & 0 & 1 & 2 & 1 \\ 1 & 1 & 1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 2 & 1 & 0 \\ 0 & 2 & 2 & 1 & 0 & 1 \end{pmatrix}$ and $M^P \oplus M^Q - M^P \otimes M^Q = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$, from Table 1.

2.3. The Similarity Measure Matrix

In information systems, the division $U/IND(R)$ determined by R is also called a granular structure. From the perspective of granular computing, neither information entropy nor knowledge granularity is able to measure the differences between granular structures [33]. Equivalence is not implied for one granular structure and another with the same granularity, i.e., granularity cannot portray the differences [34] or similarities between granular structures.

For example, if $P = \{Price\}$ and $Q = \{Size\}$, then $GD(P) = GD(Q) = 1/2$, but $IND(P) \neq IND(Q)$.

The distance of knowledge was proposed in the literature to measure the difference between granular structures [34,35]. The similarity measure is defined between information granules based on set similarity [35], and the similarity measure matrix for granular structures is constructed in this section.

In $K = (U, AT)$, suppose that $M^P = \left(\overrightarrow{X_1^P}; \overrightarrow{X_2^P}; \cdots; \overrightarrow{X_m^P}\right)$ and $M^Q = \left(\overrightarrow{X_1^Q}; \overrightarrow{X_2^Q}; \cdots; \overrightarrow{X_n^Q}\right)$ are granular matrices induced by $P, Q (P, Q \subseteq AT)$, respectively, where $\overrightarrow{X_i^P} = \left(x_{i1}^P, x_{i1}^P, \cdots, x_{i|U|}^P\right)$ and $\overrightarrow{X_j^Q} = \left(x_{j1}^Q, x_{j1}^Q, \cdots, x_{j|U|}^Q\right), 1 \leq i \leq m \leq |U|, 1 \leq j \leq n \leq |U|$.

The similarity measure between information granules and the similarity measure matrix for granular structures are defined as follows.

Definition 7. $s\left(X_i^P, X_j^Q\right) = \left|X_i^P \cap X_j^Q\right| / \left|X_i^P \cup X_j^Q\right|$ *is called the similarity between* X_i^P *and* X_j^Q. *If the row vectors of the information granules* X_i^P *and* X_j^Q *are* $\overrightarrow{X_i^P} = \left(x_{i1}^P, \cdots, x_{i|U|}^P\right)$, $\overrightarrow{X_i^Q} = \left(x_{i1}^Q, \cdots, x_{i|U|}^Q\right)$, *respectively, then the similarity is*

$$s\left(X_i^P, X_j^Q\right) = \frac{\left|\overrightarrow{X_i^P} \otimes \overrightarrow{X_j^Q}\right|}{\left|\overrightarrow{X_i^P} + \overrightarrow{X_j^Q} - \overrightarrow{X_i^P} \otimes \overrightarrow{X_j^Q}\right|} = \frac{\sum_{k=1}^{|U|} x_{ik}^P \cdot x_{jk}^Q}{\sum_{k=1}^{|U|}\left(x_{ik}^P + x_{jk}^Q - x_{ik}^P \cdot x_{jk}^Q\right)}.$$

If $\overrightarrow{X_i^P} = \overrightarrow{X_j^Q}$, then $s\left(X_i^P, X_j^Q\right) = 1$. If $\overrightarrow{X_i^P} \cdot \overrightarrow{X_j^Q} = 0$, then $s\left(X_i^P, X_j^Q\right) = 0$. In particular, if $\forall X_i^P, X_j^P \subseteq U/IND(P)$, then $s\left(X_i^P, X_j^P\right) = 0, i \neq j$.

Definition 8. $M^{s(P \cdot Q)} = \left(s_1(\overrightarrow{P} \cdot Q); \cdots; s_m(\overrightarrow{P} \cdot Q)\right)$ *is called the similarity measure matrix between* $U/IND(P)$ *and* $U/IND(Q)$, *where* $s_i(\overrightarrow{P}, Q) = \left(s\left(X_i^P, X_1^Q\right), \cdots, s\left(X_i^P, X_n^Q\right)\right)$.

For example, if $P = \{Price\}$ and $Q = \{Size\}$, then $\overrightarrow{X_1^P} = (1,0,0,0,1,1)$, $\overrightarrow{X_2^P} = (0,1,1,1,0,0)$, $\overrightarrow{X_1^Q} = (1,0,0,1,1,0)$, and $\overrightarrow{X_2^Q} = (0,1,1,0,0,1)$. Based on Definition 7, $s\left(X_1^P, X_1^Q\right) = 1/2$, $s\left(X_1^P, X_2^Q\right) = 1/5$, $s\left(X_2^P, X_1^Q\right) = 1/5$, and $s\left(X_2^P, X_2^Q\right) = 1/2$. Based on Definition 8, $M^{s(P \cdot Q)} = \begin{pmatrix} 1/2 & 1/5 \\ 1/5 & 1/2 \end{pmatrix}$, from Table 1.

In particular, if $U/IND(P) = U/IND(Q)$, then $M^P = M^Q$, and the similarity measure matrix $M^{s(P \cdot Q)} = I$ is the unit matrix.

The similarity measure matrix portrays the correlation between attribute divisions. The element at each position in the matrix presents the degree of similarity between information granules.

3. A Decision Tree Classification Algorithm Based on Granular Matrices

The decision tree classification algorithm based on granular matrices (GMDT) is a new decision tree construction method. In this algorithm, the classification accuracy of conditional attributes is used to select the classified attributes as the nodes of the tree, and it is calculated by weighting classification actions using the probability distribution determined by conditional attributes.

3.1. Selection Criteria of Classification Attributes

The key technique of a decision tree classification algorithm is how to select the appropriate splitting attribute to determine the nodes of the tree. The classification accuracy of conditional attributes is defined and used to select the splitting attribute.

In $K = (U, C \cup D)$, suppose that $M^{s(P \cdot D)} = \left(s_1(\overrightarrow{P} \cdot D); \cdots; s_m(\overrightarrow{P} \cdot D)\right)$ is the similarity measure matrix between $U/IND(P), P \subseteq C$ and $U/IND(D)$, where $s_i(\overrightarrow{P}, D) = \left(s\left(X_i^P \cdot X_1^D\right), \cdots, s\left(X_i^P \cdot X_n^D\right)\right)$.

The classification accuracy of conditional attributes is then defined as follows.

Definition 9. *In* $K = (U, C \cup D)$, *if the mapping* $s_i(P \cdot D) = max\left\{s\left(X_i^P \cdot X_1^D\right), \cdots, s\left(X_i^P \cdot X_n^D\right)\right\}$ *is defined in* $s_i(\overrightarrow{P} \cdot D)$, *then* $a(P \cdot D) = (s_1(P \cdot D), s_1(P \cdot D), \ldots, s_m(P \cdot D))$ *is called the classifica-*

tion action of P, and $\alpha^{(P,D)} = \sum_{i=1}^{m} P_i^P \cdot a_i(P \cdot D)$ *is called the classification accuracy of P relative to the decision attributes D.*

For example, if $P = \{Price\}$ and $D = \{Buy\}$, then $\vec{X_1^P} = (1,0,0,0,1,1)$, $\vec{X_2^P} = (0,1,1,1,0,0)$, $\vec{X_1^D} = (1,0,1,0,0,1)$, and $\vec{X_2^D} = (0,1,0,1,1,0)$. Based on Definition 7, $s(X_1^P, X_1^D) = 1/2$, $s(X_1^P, X_2^D) = 1/5$, $s(X_2^P, X_1^D) = 1/5$, $s(X_2^P, X_2^D) = 1/2$, and $P_i^P = |X_i^P|/|U| = 1/2, i = 1,2$. Based on Definition 8, $M^{s(P \cdot D)} = \begin{pmatrix} 1/2 & 1/5 \\ 1/5 & 1/2 \end{pmatrix}$. Based on Definition 9, the classification action of P is $a(P \cdot D) = (1/2, 1/2)$, and the classification accuracy of P relative to the decision attributes D is $\alpha^{(P,D)} = 1/2$, from Table 1.

The classification accuracy $\alpha^{(P,D)}$ of P relative to D portrays the cognitive ability of $U/IND(D)$ based on $U/IND(P)$. In particular, $\alpha^{(P,D)} = 1$ shows that the cognition (division) of $U/IND(P)$ to $U/IND(D)$ is identical. When the splitting attribute of the decision tree is selected according to $\alpha^{(P,D)}$, a splitting attribute highly consistent with the classification result of the decision attribute can be extracted from the many classification attributes as the node of the tree, and the classification accuracy of the decision tree can thus be improved.

In addition, $GD(P) = 1$ indicates that the attribute has no classification ability and cannot be used as a splitting attribute of the decision tree.

3.2. Selection Method of Leaf Nodes

In the decision tree branching process, the dataset is divided into multiple subsets by the splitting attribute, and the branching process is recursively invoked for each subset until each subset contains only the same type of data, at which point the classification ends. In granular computing, the elements contained in each information granule are indistinguishable; that is to say, the data types are consistent. In GMDT, if $s\left(X_i^P, X_j^D\right) = \left|X_i^P \cap X_j^D\right| / \left|X_i^P \cup X_j^D\right| = |X_i^P| / \left|X_j^D\right|$ between X_i^P and X_j^D, then $X_i^P \subseteq X_j^D$. This indicates that the conditional attribute cannot further subdivide the decision granule X_j^D, and X_i^P will be made a leaf node of the decision tree.

3.3. Algorithm Description and Analysis

The core idea of GMDT is to select the conditional attribute with the highest classification accuracy as the node of the tree through the operation of the granular matrix; the branching process is recursively invoked for each subset until each subset contains only the same type of data, at which point the classification ends. The specific algorithm flow is shown in Algorithm 1.

3.4. Time and Space Complexity

Time complexity: Assuming that the data-set has n samples and m attributes, in phase 1, the computational effort of the algorithm is mainly to calculate the granular matrix induced by each attribute, which produces a time complexity of $O(m \times n^2)$. Next, according to the granular matrices induced by the attributes, bit-multiplication and bit-sum operations are performed to calculate the similarity matrix between each information grain and the decision grain in the categorical attributes. Assuming that there are k classification attributes, the time complexity is $O(k \times n^2)$. Finally, in the classification accuracy function, for each attribute, its classification accuracy must be calculated. Assuming that there are n data samples and m attributes, the time complexity of computing the classification accuracy for each attribute is $O(n \times m)$. In the second stage, the amount of calculation is mainly due to recursive calls. If the data-set is not empty, the classification accuracy function and the whole algorithm are called recursively. Assuming that the size of the data-set is p and the recursive depth of the algorithm is d, the time complexity of the recursive calls is $O(p \times d)$. Thus, the time complexity of this algorithm is $O(m \times n^2)$.

Space complexity: The process of recursively calling the classification accuracy function and the whole process of the algorithm involve the use of stack space. Assuming that the recursion depth is d, each recursive call needs to save the local variables and parameters of the function, so the space complexity of this algorithm is O(d).

Algorithm 1: Decision Tree Algorithm Based on Granular Matrices

Input: K = (Data: the set of training samples, Attributes: the set of conditional and decision attributes)
Output: Tree

```
1: Classification accuracy(k,data,attributes)
2: if k >= len(attributes)
        return
3: if Granular Degree(data, conditional attributes, k) == 1
     Classification accuracy[k] = 0
   else
     Granular Matrix (Data, Attributes)
     Similar Matrix = Granular Matrix (conditional) ⊗ Granular Matrix(decision)
     Classification Action = max (Similar Matrix, axis = 1)
     Classification accuracy[k] = Classification Action*Granular Degree
     Classification accuracy (k + 1,data,attributes)
4: select Node = max (Classification accuracy)
5: Tree = Tree + Node
6: update (data, attributes), Classification accuracy = {}
7: if data! = null
8:    go to 1
9: else
10:   return tree
```

3.5. Calculation Example

The decision information system $K = (U, C \cup D)$ is shown in Table 2.

Table 2. A decision information system, K = (U,C∪D).

U	C						D
	Color	**Root**	**Stroke**	**Texture**	**Umbilical**	**Touch**	**Good**
1	D-green	Curl up	Turbid	Clear	Depressed	Smooth	Yes
2	Jet-black	Curl up	Dull	Clear	Depressed	Smooth	Yes
3	Jet-black	Curl up	Turbid	Clear	Depressed	Smooth	Yes
4	D-green	Curl up	Dull	Clear	Depressed	Smooth	Yes
5	Plain	Curl up	Turbid	Clear	Depressed	Smooth	Yes
6	D-green	Slightly curl	Turbid	Clear	Concave	Soft	Yes
7	Jet-black	Slightly curl	Turbid	S-vague	Concave	Soft	Yes
8	Jet-black	Slightly curl	Turbid	Clear	Concave	Smooth	Yes
9	Jet-black	Slightly curl	Dull	S-vague	Concave	Smooth	No
10	D-green	Stiff	Melodious	Clear	Flat	Soft	No
11	Plain	Stiff	Melodious	Vague	Flat	Smooth	No
12	Plain	Curl up	Turbid	Vague	Flat	Soft	No
13	D-green	Slightly curl	Turbid	S-vague	Depressed	Smooth	No
14	Plain	Slightly curl	Dull	S-vague	Depressed	Smooth	No
15	Jet-black	Slightly curl	Turbid	Clear	Concave	Soft	No
16	Plain	Curl up	Turbid	Vague	Flat	Smooth	No
17	D-green	Curl up	Dull	S-vague	Concave	Smooth	No

Based on Definition 7, the similarity measure matrices of M^c, M^r, M^s, M^{te}, M^u, and M^D are

$$M^{s(c,D)} = \begin{pmatrix} 3/11 & 3/12 \\ 4/10 & 2/13 \\ 1/12 & 4/10 \end{pmatrix}, M^{s(r,D)} = \begin{pmatrix} 5/11 & 3/14 \\ 3/12 & 4/12 \\ 0 & 2/9 \end{pmatrix}, M^{s(s,D)} = \begin{pmatrix} 6/12 & 4/15 \\ 2/11 & 3/11 \\ 0 & 2/9 \end{pmatrix}$$

$$M^{s(te,D)} = \begin{pmatrix} 7/10 & 2/16 \\ 1/12 & 4/10 \\ 0 & 3/9 \end{pmatrix}, M^{s(u,D)} = \begin{pmatrix} 5/10 & 2/14 \\ 3/11 & 3/12 \\ 0 & 4/9 \end{pmatrix}, M^{s(to,D)} = \begin{pmatrix} 6/14 & 6/15 \\ 2/11 & 3/11 \end{pmatrix}$$

The classification decision actions are as follows:

$$a(c,D) = \left(\frac{3}{11}, \frac{4}{10}, \frac{4}{10}\right)' a(r,D) = \left(\frac{5}{11}, \frac{4}{12}, \frac{2}{9}\right)' a(s,D) = \left(\frac{6}{12}, \frac{3}{11}, \frac{2}{9}\right)$$

$$a(te,D) = \left(\frac{7}{10}, \frac{4}{10}, \frac{3}{9}\right)' a(u,D) = \left(\frac{5}{10}, \frac{3}{11}, \frac{4}{9}\right)' a(\ to\ ,D) = \left(\frac{6}{14}, \frac{3}{11}\right)$$

The classification accuracies of Color, Root, Stroke, Texture, Umbilical, and Touch relative to D are $\alpha^{(c,D)} = 0.3551$, $\alpha^{(r,D)} = 0.3773$, $\alpha^{(s,D)} = 0.4005$, $\alpha^{(te,D)} = 0.5471$, $\alpha^{(u,D)} = 0.4067$, and $\alpha^{(to,D)} = 0.3827$.

Obviously, $\alpha^{(te,D)} = 0.5471$ is the highest value, so Texture is selected as the splitting attribute to form a node in the tree to divide the data-set.

Similarly, the classification accuracies of the conditional attributes in each sub-dataset are calculated, and the attribute with the highest accuracy is selected as the division attribute for recursion; the constructed decision tree is shown in Figure 1.

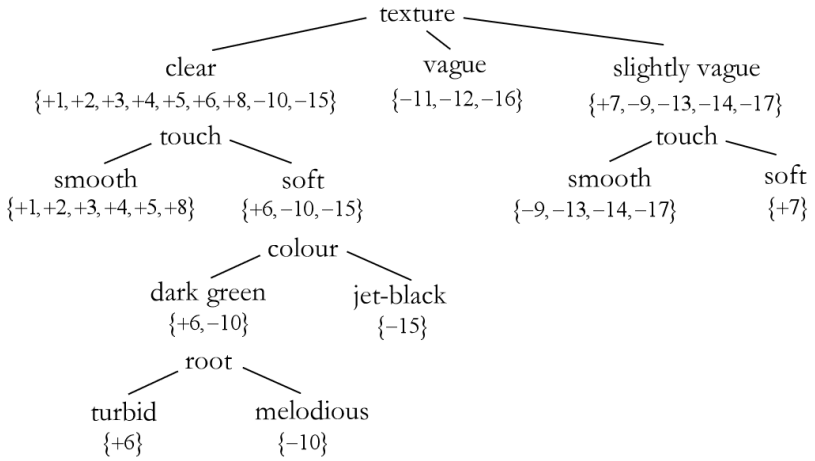

Figure 1. Decision tree constructed by GMDT.

4. Experiments and Result Analysis

4.1. Experimental Environment

The computer hardware and environment configuration were an AMD Ryzen 75800H 3.20 GHz processor with Radeon Graphics, 16.0 GB of RAM, a Windows 10 operating system, and Python version 3.10.

Many existing decision tree models were developed and optimized by combining other theories based on commonly used operators such as the information gain, Taylor's formula, the Gini coefficient, granular computing, etc. The basis of these decision tree models is these commonly used operators. The decision tree classification algorithm based on granular matrices proposes a new splitting attribute selection operator, i.e., the classification accuracy of the splitting attribute relative to the decision attribute, changing the core of the classification algorithm. Therefore, the classical ID3 and C4.5 were selected for comparison with GMDT in this paper.

In our experiments, more than 20 sets of classification data were selected from UCI through a "classification" keyword search to test the effectiveness of the classification

algorithm, including Iris, Heart Disease, Breast Cancer, Adult, Win, CRX, Wine Quality, Car Evaluation, Acute Inflammations, Krkogt, dermatology, etc. In the actual situation where these experimental data were not preprocessed, the classification accuracy of the GMDT, ID3, and C4.5 algorithms on all but the Acute Inflammations, Iris, Breast Cancer, CRX, Mushrooms, and Adult datasets was less than 60%. Therefore, only the experimental results on datasets with classification accuracy higher than 60%, as shown in Table 3, are presented.

Table 3. Description of UCI datasets.

No.	Dataset	Samples	Attributes	Class
1	Acute Inflammations	121	6	2
2	Iris	151	4	3
3	Breast Cancer	286	9	2
4	CRX	654	15	2
5	Mushrooms	8125	21	2
6	Adult	32,561	14	2

4.2. Algorithm Comparison Experiment

(1) Comparison of classification accuracy

The six experimental datasets in Table 2 were selected and partitioned into 70% for the training set and 30% for the test set. ID3, C4.5, and GMDT were run. The classification accuracies calculated in the experiment are shown in Table 4.

Table 4. Classification accuracy.

No.	Data	Classification Accuracy		
		ID3	C4.5	GMDT
1	Acute Inflammations	0.64	0.64	0.64
2	Iris	0.80	0.87	0.90
3	Breast Cancer	0.46	0.53	0.61
4	CRX	0.31	0.34	0.63
5	Mushrooms	0.80	0.87	0.90
6	Adult	0.27	0.45	0.75

In order to intuitively present the experimental results, "blue" was chosen to represent the experimental calculation results of ID3, "red" was chosen to represent the experimental calculation results of C4.5, and "orange" was chosen to represent the experimental calculation results of GMDT. A histogram of the experimental results of the three classification algorithms was drawn, as shown in Figure 2.

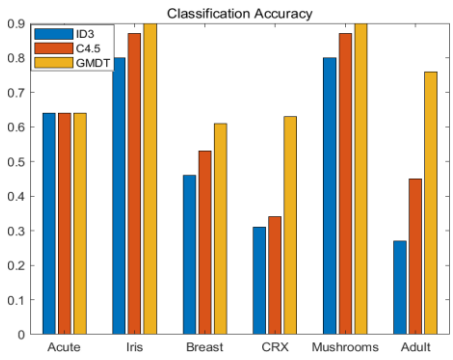

Figure 2. Comparison chart of classification accuracy.

The following can be clearly observed: ① The accuracy of the GMDT classification algorithm was higher than that of the classical ID3 and C4.5, especially on the CRX and Adult datasets. ② The classification accuracies of all three algorithms were affected by the data type, the classification accuracies for different data types were different, and the classification accuracies of the three algorithms were basically at the same level in the same data. The maximum variance of the classification accuracy was only 0.0392, as shown in Figure 3, which indicates that the robustness of the GMDT decision tree classification algorithm and that of the classical ID3 and C4.5 classification algorithms are basically consistent.

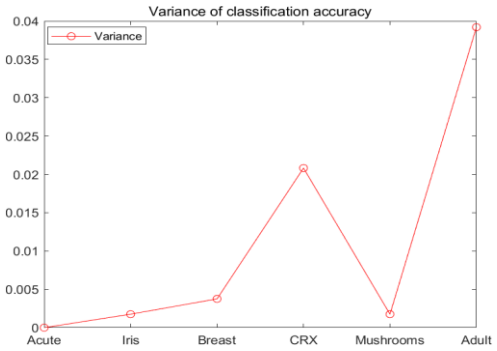

Figure 3. Variance of classification accuracy.

(2) Error analysis

To further compare and analyze the classification performance of ID3, C4.5, and GMDT, the numbers of correctly classified, unclassified, and misclassified samples in the test set were extracted from the experimental calculation results and expressed as a ternary ordered array (correct classification, no classification, and classification error), as shown in Table 5.

Table 5. Error analysis.

No.	Data	Error Analysis		
		ID3	C4.5	GMDT
1	Acute Inflammations	(23, 0, 13)	(23, 0, 13)	(23, 0, 13)
2	Iris	(36, 7, 2)	(39, 3, 3)	(41, 3, 1)
3	Breast Cancer	(35, 10, 31)	(40, 6, 30)	(46, 0, 30)
4	CRX	(64, 110, 34)	(70, 101, 37)	(131, 20, 57)
5	Mushrooms	(1950, 340, 148)	(2126, 168, 144)	(2197, 0, 241)
6	Adult	(2679, 5937, 1153)	(4396, 4662, 738)	(7352, 33, 2384)

In order to intuitively present the experimental results, "blue" was chosen to represent the experimental calculation results of ID3, "red" was chosen to represent the experimental calculation results of C4.5, and "orange" was chosen to represent the experimental calculation results of GMDT. Histograms of the experimental results from the three classification algorithms in terms of the rate of unclassified data, the rate of classification errors, and the classification potential are plotted in Figure 4, Figure 5, and Figure 6, respectively.

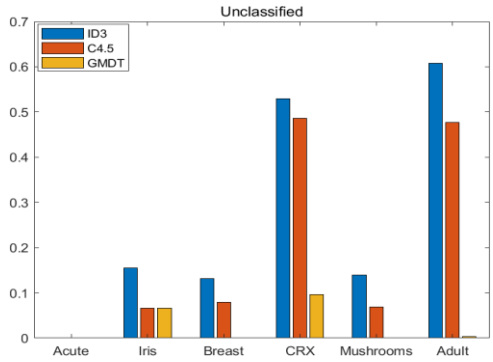

Figure 4. Variance of unclassified data.

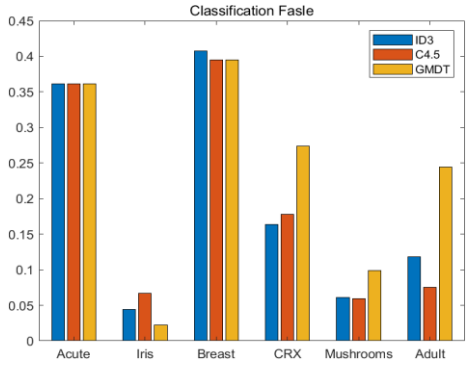

Figure 5. Comparison chart of classification errors.

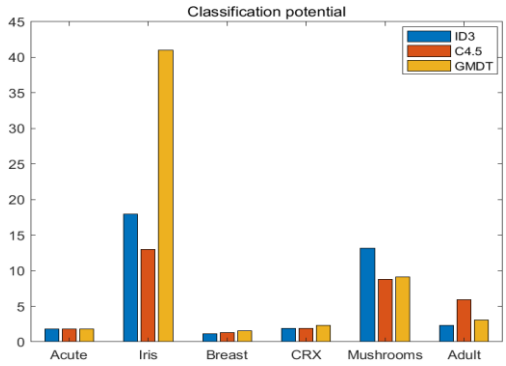

Figure 6. Comparison chart of classification potential.

Since the model training of the GMDT classification algorithm requires the classification results of split attributes to be highly consistent with the decision classification, the completed training decision tree classification model showed slight overfitting in the test set. The specific performance was as follows: ① The unclassified proportion of the GMDT classification algorithm was much smaller than those of ID3 and C4.5. As shown in Figure 4, the unclassified ratio of GMDT was 0.096 in the CRX dataset, while the unclassified ratios of ID3 and C4.5 were as high as 0.53 and 0.49, respectively. In the Adult dataset, the unclassified ratio of GMDT was only 0.0034, while the unclassified ratios of ID3 and C4.5 were as high as 0.61 and 0.48, respectively. The difference became more prominent as the sample size of the dataset increased. ② The classification error rate of the GMDT algorithm increased with increasing data volume and data diversity. As shown in Figure 5, the classification error rate of the GMDT classification algorithm was greater than those of ID3 and C4.5 in both the CRX and Adult datasets.

(3) Comparison of classification potential

The set-pair potential [36] is used to portray the developmental trend of a system by comparing the same degree with the opposition degree. In this paper, the classification accuracy was taken as the same degree, while the error rate was regarded as the opposition degree; the classification potential of ID3, C4.5, and GMDT was thus compared. As shown in Figure 6: the classification potential was basically consistent, even though the classification error rate of GMDT was higher than those of ID3 and C4.5 in the two large-sample datasets, CRX and Adult, which indicates that the slight overfitting did not affect GMDT's classification performance. GMDT's classification potential in the Adult dataset was much higher than that of ID3 and C4.5, which indicates that GMDT has better classification performance than ID3 and C4.5.

5. Conclusions

In this paper, a decision tree classification algorithm based on granular matrices was proposed and introduced in detail. Experiments were performed to compare GMDT and the classical ID3 and C4.5; the test results showed that the proposed algorithm has higher classification accuracy and better classification performance than ID3 and C4.5. These studies further optimized the classification performance and extended the classification method. Further research is needed on the operational laws of granular matrices; research following on from this paper will focus on constructing multivariate decision trees and random forests based on similarity granular matrices.

Author Contributions: Conceptualization, L.M. and B.B.; formal analysis, W.Z. and C.Z.; funding acquisition, C.Z. and L.L. All authors have read and agreed to the published version of the manuscript.

Funding: The Hebei Province Professional Degree Teaching Case Establishment and Construction Project (Chunying Zhang: No. KCJSZ2022073).

Data Availability Statement: The data used for the experimental analysis came from the public UCI datasets mentioned in the article.

Acknowledgments: Support from colleagues and the university is acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fu, C. Research on Data Classification Algorithm Based on Granular Computing. Ph.D. Thesis, Dalian University of Technology, Dalian, China, 2021.
2. Zhang, C.L.; Zhang, L. A New ID3 Algorithm Based on Revised Information Gain. *Comput. Eng. Sci.* **2008**, *30*, 46–47.
3. Jin, C.; Luo, D.-L.; Mu, F.-X. An improved ID3 decision tree algorithm. In Proceedings of the 2009 4th International Conference on Computer Science & Education, Nanning, China, 25–28 July 2009; pp. 127–130.
4. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: Oxford, UK, 2017.
5. Prasad, L.V.; Naidu, M.M. CC-SLIQ: Performance Enhancement with 2k Split Points in SLIQ Decision Tree Algorithm. *IAENG Int. J. Comput. Sci.* **2014**, *41*, 163–173.
6. Wei, H.N. Research on Parallel Decision Tree Classification Based on SPRINT Method. *Comput. Appl.* **2005**, 39–41.
7. Chandra, B.; Varghese, P.P. Fuzzy SLIQ decision tree algorithm. *IEEE Trans. Syst. Man Cybern. Part B* **2008**, *38*, 1294–1301. [CrossRef]
8. Honglei, G.; Changqian, M.; Wenjian, W. Model decision tree algorithm for multicore Bayesian optimization. *J. Natl. Univ. Def. Technol.* **2022**, *44*, 67–76.
9. Yun, J.; Seo, J.W.; Yoon, T. Fuzzy decision tree. *Int. J. Fuzzy Log. Syst.* **2014**, *4*. [CrossRef]
10. Bujnowski, P.; Szmidt, E.; Kacprzyk, J. An approach to intuitionistic fuzzy decision trees. In Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15), Asturias, Spain, 30 June–3 July 2015; pp. 1253–1260.
11. Yu, B.; Guo, L.; Li, Q. A characterization of novel rough fuzzy sets of information systems and their application in decision making. *Expert Syst. Appl.* **2019**, *122*, 253–261. [CrossRef]
12. Marudi, M.; Ben-Gal, I.; Singer, G. A decision tree-based method for ordinal classification problems. *IISE Trans.* **2022**, 1–15. [CrossRef]
13. Wang, Y. An ordered decision tree algorithm based on fuzzy dominant complementary mutual information. *Comput. Appl.* **2021**, *41*, 2785–2792.

14. Zhao, H.; Wang, P.; Hu, Q.; Zhu, P.; Xu, J.; Fang, H.; Zhou, T.; Chen, Y.-H.; Guo, H.; Zeng, F. Fuzzy rough set based feature selection for large-scale hierarchical classification. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 1891–1903. [CrossRef]
15. Tawhid, M.A.; Ibrahim, A.M. Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 573–602. [CrossRef]
16. Yue, Y.; Xian, Z.; Shuai, C. Decision tree induction algorithm based on attribute purity. *Comput. Eng. Des.* **2021**, *42*, 142–149.
17. Wang, R.; Liu, Z.; Ji, J. Decision tree algorithm based on attribute importance. *Comput. Sci.* **2017**, *44*, 129–132.
18. Xie, X.; Zhang, X.Y.; Yang, J.L. A decision tree algorithm incorporating information gain and Gini index. *Comput. Eng. Usage* **2022**, *58*, 139–144.
19. Wu, S.B.; Chen, C.G.; Huang, R. ID3 optimization algorithm based on correlation coefficient. *Comput. Eng. Sci.* **2016**, *38*, 2342–2347.
20. Choi, S.H.; Shin, J.M.; Choi, Y.H. Dynamic nonparametric random forest using covariance. *Secur. Commun. Netw.* **2019**, *2019*, 1–12. [CrossRef]
21. Kozak, J. *Decision Tree and Ensemble Learning Based on Ant Colony Optimization*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019.
22. Mu, Y.; Liu, X.; Yang, Z.; Liu, X. A parallel C4. 5 decision tree algorithm based on MapReduce. *Concurr. Comput. Pract. Exp.* **2017**, *29*, e4015. [CrossRef]
23. Liang, J.; Qian, Y.; Li, D.; Hu, Q. Research progress on theory and method of granular computing for big data. *Big Data* **2016**, *2*, 13–23.
24. Miao, D.; Hu, S. Uncertainty analysis based on granular computing. *J. Northwest Univ.* **2019**, *49*, 487–495.
25. Khazali, N.; Sharifi, M.; Ahmadi, M.A. Application of fuzzy decision tree in EOR screening assessment. *J. Pet. Sci. Eng.* **2019**, *177*, 167–180. [CrossRef]
26. Rabcan, J.; Rusnak, P.; Kostolny, J.; Stankovic, R. Comparison of algorithms for fuzzy decision tree induction. In Proceedings of the 2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), Kosice, Slovenia, 12–13 November 2020; pp. 544–551.
27. Zhang, K.; Zhan, J.; Wu, W.Z. Novel fuzzy rough set models and corresponding applications to multi-criteria decision-making. *Fuzzy Sets Syst.* **2020**, *383*, 92–126. [CrossRef]
28. Qian, W.; Huang, J.; Wang, Y.; Xie, Y. Label distribution feature selection for multi-label classification with rough set. *Int. J. Approx. Reason.* **2021**, *128*, 32–55. [CrossRef]
29. Chen, C.E.; Ma, X. Information system rule extraction method based on granular computing. *J. Northwest Norm. Univ.* **2018**, *54*, 11–15.
30. Wu, J.; Wang, C. Multidimensional granular matrix correlation analysis for big data and applications. *Comput. Sci.* **2017**, *44*, 407–410+421.
31. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356. [CrossRef]
32. Miao, Z.; Li, D. *Rough Set Theory Algorithms and Applications*; Tsinghua University Press: Beijing, China, 2008.
33. Yang, J.; Wang, G.; Zhang, Q. Evaluation model of rough grain structure based on knowledge distance. *J. Intell. Syst.* **2020**, *15*, 166–174.
34. Yang, J.; Wang, G.; Zhang, Q. Similarity metrics for multi-granularity cloud models. *Pattern Recognit. Artif. Intell.* **2018**, *31*, 677–692.
35. Qian, Y.; Liang, J.; Dang, C. Knowledge structure, knowledge granulation and knowledge distance in a knowledge base. *Int. J. Approx. Reason.* **2009**, *50*, 174–188. [CrossRef]
36. Zhou, R.; Jin, J.; Cui, Y.; Ning, S.; Bai, X.; Zhang, L.; Zhou, Y.; Wu, C.; Tong, F. Agricultural drought vulnerability assessment and diagnosis based on entropy fuzzy pattern recognition and subtraction set pair potential. *Alex. Eng. J.* **2022**, *61*, 51–63. [CrossRef]