

Article



# Multi-Label Attribute Reduction Based on Neighborhood Multi-Target Rough Sets

Wenbin Zheng <sup>1,2,\*</sup>, Jinjin Li<sup>3</sup>, Shujiao Liao <sup>3,\*</sup>, and Yidong Lin<sup>3</sup>

- <sup>1</sup> School of Computer Science, Minnan Normal University, Zhangzhou 363000, China
- <sup>2</sup> Key Laboratory of Data Science and Intelligence Application, Fujian Province University, Zhangzhou 363000, China
- <sup>3</sup> School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China
- \* Correspondence: zwb2077@mnnu.edu.cn (W.Z.); sjliao2011@163.com (S.L.)

Abstract: The rough set model has two symmetry approximations called upper approximation and lower approximation, which correspond to a concept's intension and extension, respectively. Multilabel learning enforces the rough set model, which wants to be applied considering the correlations among labels, while the target concept should not be limited to only one. This paper proposes a multi-target model considering label correlation (Neighborhood Multi-Target Rough Sets, NMTRS) and proposes an attribute reduction approach based on NMTRS. First, some definitions of NMTRS are introduced. Second, some properties of NMTRS are discussed. Third, some discussion about the attribute significance measure is given. Fourth, the attribute reduction approaches based on NMTRS are proposed. Finally, the efficiency and validity of the designed algorithms are verified by experiments. The experiments show that our algorithm shows considerable performance when compared to state-of-the-art approaches.

Keywords: multi-label learning; attribute reduction; multi-target rough set; label correlation



Citation: Zheng, W.; Li, J.; Liao, S.; Lin, Y. Multi-Label Attribute Reduction Based on Neighborhood Multi-Target Rough Sets. *Symmetry* 2022, 14, 1652. https://doi.org/ 10.3390/sym14081652

Academic Editor: László T. Kóczy

Received: 8 July 2022 Accepted: 4 August 2022 Published: 10 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). 1. Introduction

Since rough set theory was proposed by Pawlak [1] in 1982, it quickly became a hot topic in knowledge discovery and has been widely used in many applications such as classification [2–6], clustering [7–11], and attribute reduction [12]. It has two approximations corresponding to a target concept's intension and extension, which shows symmetry. Various rough set models are based on different types of binary relations, such as multiple equivalence relations [13], general binary relations [14], and so on. Within these models, the neighborhood relation is outstanding for its ability to deal with both nominal and numerical attributes at the same time.

There are lots of works applying the neighborhood rough set model in various fields. Inbarani et al. [15] proposed a classification algorithm by using the neighborhood rough set model. For dynamic data mining, Zhang et al. [16] proposed a neighborhood rough set approach. Most relevant works applied the neighborhood rough set model to attribute reduction tasks.

Attribute reduction, or feature selection, is a traditional but essential machine learning task. Attribute reduction approaches try to select some features from the raw attribute set without harming the data's information presentation ability. These approaches have made remarkable achievements in eliminating noise and promoting learning time efficiency. For attribute reduction tasks, there are different types of work. The first type is single-label attribute reduction. Hu et al. [17] proposed an approach for attribute reduction based on neighborhood rough sets. A quick attribute reduction algorithm was proposed by Yong et al. [18] based on the neighborhood rough set model. Additionally, there are parallel attribute reduction approaches [19], online streaming attribute reduction [20], and so on.

These attribute reduction approaches were proposed based on the classic or extended neighborhood rough set model.

The second type of attribute reduction task is multi-label attribute reduction. These methods use strategies to handle the multiple labels in a multi-label learning paradigm. For example, Sun et al. [21] proposed a multi-label attribute reduction approach by transforming the multi-label learning problem into a single label one by ignoring the correlations within labels. Fisher R.A. et al. [22] regard the feature space and the label space as two different viewpoints of the data to improve the original dimension reduction method. By using the kernel matrix method, Wold H. et al. [23] proposed a similar method to [22]. By using the mapping dimensionality reduction and sub-control dimensionality reduction approaches at the same time, Zhang et al. [24] proposed a dimensionality reduction with a linear kernel matrix or non-linear kernel matrix. Based on PCA and genetic algorithms, Zhang et al. [25] proposed MLNB by using the Naive Bayes method to extract the features simultaneously. Liu et al. [26] proposed an attribute reduction method based on a neighborhood rough set model. Meanwhile, an online multi-label attribute reduction method was proposed by Liu et al. [27] using the neighborhood rough set model. The f-neighborhood rough set model was used to derive the feature selection method for multi-label learning [28].

The attribute reduction methods designed for single-label learning or multi-label learning based on neighborhood rough sets all use the classic neighborhood rough set model and its extended models. All of these neighborhood rough set models are designed for a classic information system [29–35], which means that the system only has one decision attribute. None of them are designed for a multi-decision information system that considers label correlation among labels simultaneously.

In this paper, we propose a neighborhood multi-target rough set model and then design an attribute reduction algorithm based on it. We design the model by defining a global correlated target set to be the target group of the rough set model. The coefficient of the global correlated target set controls the relevance of different target concepts. Then, we use a conservative strategy to combine the correlated target to define the rough set model. Using the rough set model which we proposed, the attribute significance measure can be then given. Furthermore, we can derive the corresponding attribute reduction algorithm.

The contributions of this paper are as follows:

- A neighborhood rough set model considering the label correlation is proposed for multi-label learning.
- The properties of the proposed models are investigated.
- An algorithm for calculating the approximations in the proposed rough set model is designed.
- Attribute significance measure is given based on the rough set model we proposed.
- Experiments are conducted to validate the efficiency and effectiveness of the proposed algorithms.

The rest of this paper is organized as follows. Some basic concepts of NMTRS are introduced, and their properties are discussed in Section 2. In Section 3, the attribute significance measure is given, along with some discussions about it, and the corresponding attribute reduction algorithms are derived by the significance measure. All the algorithms are evaluated in Section 4. Finally, we conclude the whole paper in Section 5.

#### 2. Neighborhood Multi-Target Rough Sets

In this section, some concepts associated with our proposed model are introduced and then the properties of the proposed model are discussed.

#### 2.1. Definitions

In this subsection, the definitions of neighborhood multi-target rough sets are introduced.

**Definition 1.** [1] Suppose U is a finite universe and  $A = \{a_1, a_2, \dots, a_m\}$  is an attribute set, then (U, A) is an information system.

**Definition 2.** [35] (Set Correlation, SC) Suppose U is a finite universe and (U, A) is an information system,  $X = \{X_1, X_2, \dots, X_r\}$  is a finite target set which satisfies for all  $X_i \in X, X_i \subseteq U$ . The set *correlation is defined by:* 

$$S_R(X_i, x_j) = |X_i \cap X_j| / |X_i|, i, j \le r \text{ (Relative Correlation) or}$$
$$S_R(X_i, x_j) = |X_i \cap X_j| / |X_i \cup X_j|, i, j \le r \text{ (Absolute Correlation).}$$

**Definition 3.** (Global Correlated Target Set, GCTS) Suppose U is a finite universe and (U, A)is an information system,  $\mathbf{X} = \{X_1, X_2, \dots, X_r\}$  is a finite target set which satisfies for all  $X_i \in X, X_i \subseteq U$ . Then, X is a global correlated target set if and only if for all  $X_i, X_i \in X$ ,  $S_R(X_i, X_i) > \alpha$  or  $S_R(X_i, X_i) > \alpha$ ,  $\alpha \in (0, 1]$  is the Correlation Control Parameter (CCP) among targets; it controls the relevance degree among targets in the target group.

Based on the definition of GCTS and neighborhood rough sets, we can define neighborhood multi-target rough sets accordingly.

**Definition 4.** [34] (Neighborhood Class) Suppose U is a finite universe and (U, A) is an information system, where  $A = A_C \cup A_N$ ,  $A_C$  is the symbol attribute set and  $A_N$  is the numerical *attribute set. For all*  $x \in U$ ,  $\delta \ge 0$ , *the neighborhood class of* x *can be defined as* 

$$\begin{array}{l} (1) \ n_{A_{C}}(x) = \{ y \in U | \forall a \in A_{C}, y_{a} = x_{a} \}; \\ (2) \ n_{A_{N}}(x) = \{ y \in U | d(x_{A_{N}}, y_{A_{N}}) \leq \delta \}; \\ (3) \ n_{A}(x) = n_{A_{C} \cup A_{N}}(x) = \{ y \in U | \forall a \in A_{C}, y_{a} = x_{a} \wedge d(x_{A_{N}}, y_{A_{N}}) \leq \delta \} \end{array}$$

where  $y_a$  denotes the attribute value of instance y on attribute a and where  $y_{A_N}$  denotes the attribute value of instance y on attribute set  $A_N$ .

**Definition 5.** (Neighborhood Multi-Target Rough Sets, NMTRS) Suppose U is a finite universe and (U, A) is an information system,  $X = \{X_1, X_2, \dots, X_r\}$  is a finite target set which satisfies for all  $X_i \in \mathbf{X}, X_i \subseteq U$ . **X** is a GCTS and the correlation coefficient is  $\alpha$ .

Then, the lower approximation of NMTRS is defined as:

$$\underline{R}^{\alpha}_{\delta}(\mathbf{X}) = \{ x \in U | n_A(x) \subseteq X_1 \land n_A(x) \subseteq X_2 \land \dots \land n_A(x) \subseteq X_r \}$$

The upper approximation of NMTRS is defined as:

$$\overline{R}^{a}_{\delta}(\mathbf{X}) = \{ x \in U | n_{A}(x) \cap X_{1} \neq \phi \lor n_{A}(x) \cap X_{2} \neq \phi \lor \cdots \lor n_{A}(x) \cap X_{r} \neq \phi \}.$$

With the help of CCP, we can organize different targets together. The lower approximation is a conservative approximation of the target group and needs all the targets to meet the same condition. Relatively, the upper approximation obeys a liberal strategy and only needs one target of the target group to meet the condition.

**Example 1.** A multi-label decision information system is as Table 1 below: it has two labels which are assumed to be two different target concepts. We can easily obtain that when  $\alpha = 0.4$  the target group is a GCTS. For clarifying the definition of NMTRS, we have an example for NMTRS. Since

$$X_1 = \{x_2, x_3, x_5\}, X_2 = \{x_1, x_2, x_5\}, \alpha = 0.4, \delta = 0.5, S_R(X_1, X_2) = 2/3 > \alpha = 0.4.$$

From Table 1 we have

$$n_A(x_1) = \{x_1\}, n_A(x_2) = \{x_2, x_5\}, n_A(x_3) = \{x_3, x_4\}, n_A(x_4) = \{x_3, x_4\}, n_A(x_5) = \{x_2, x_5\}.$$

From Definition 5,

 $n_A(x_1) = \{x_1\} \notin X_1 \land n_A(x_1) \subseteq X_2, n_A(x_2) \subseteq X_1 \land n_A(x_2) \subseteq X_2, n_A(x_3) \notin X_1 \land n_A(x_3) \notin X_2,$  $n_A(x_4) \not\subset X_1 \wedge n_A(x_4) \not\subset X_2, n_A(x_5) \subseteq X_1 \wedge n_A(x_5) \subseteq X_2,$  $\therefore \underline{R}^{\alpha}_{\delta}(X) = \{x_2, x_5\}.$  $n_A(x_1) \cap X_2 \neq \phi, n_A(x_2) \cap X_1 \neq \phi, n_A(x_3) \cap X_1 \neq \phi, n_A(x_4) \cap X_1 \neq \phi, n_A(x_5) \cap X_1 \to \phi, n_A($ 

Table 1. An information system.

 $\therefore \overline{R}^{\alpha}_{\delta}(\mathbf{X}) = U.$ 

x	$n_1$	$a_1$	<i>a</i> <sub>2</sub>	$X_1$	$X_2$
<i>x</i> <sub>1</sub>	1.5	М	1	0	1
$x_2$	2	F	2	1	1
<i>x</i> <sub>3</sub>	2	М	2	1	0
$x_4$	1.5	М	2	0	0
<i>x</i> <sub>5</sub>	2	F	2	1	1

#### 2.2. Properties

The properties of NMTRS are discussed in this subsection.

**Proposition 1.** Suppose U is a finite universe and (U, A) is an information system,  $X = \{X_1, X_2, \cdots, X_r\}$  is a finite target set which satisfies for all  $X_i \in X, X_i \subseteq U$ . X is a GCTS and the correlation coefficient is  $\alpha$ . For the approximations of NMTRS:

(1) for all  $X_i \in \mathbf{X}, X_i = U$  if and only if  $\underline{R}^{\alpha}_{\delta}(\mathbf{X}) = U$ ; (2) exists  $X_i \in \mathbf{X}, X_i = U$  if and only if  $\overline{R}^{u}_{\delta}(\mathbf{X}) = U$ ; (3) exists  $X_i \in \mathbf{X}, X_i = \phi$  if and only if  $\underline{R}^{\alpha}_{\delta}(\mathbf{X}) = \phi$ ; (4) for all  $X_i \in \mathbf{X}, X_i = \phi$  if and only if  $\overline{R}^{\alpha}_{\delta}(\mathbf{X}) = \phi$ ; (5) for all  $X_i \in \mathbf{X}, \underline{R}^{\alpha}_{\delta}(\mathbf{X}) \subseteq X_i$ ; (6) for all  $X_i \in \mathbf{X}, X_i \subseteq \overline{R}^{\alpha}_{\delta}(\mathbf{X})$ .

# Proof.

(1) for all  $X_i \in \mathbf{X}, X_i = U$  implies for all  $x \in U$ , for all  $i \in \{1, 2, \dots, r\}, n_A(x) \subseteq X_i$ which implies  $n_A(x) \subseteq X_1 \wedge n_A(x) \subseteq X_2 \wedge \cdots \wedge n_A(x) \subseteq X_r$  so  $\underline{R}^{\alpha}_{\delta}(X) = U$ ; (2) exists  $X_i \in X$ ,  $X_i = U$  implies for all  $x \in U$ , exists  $i \in \{1, 2, \dots, r\}$ ,  $n_A(x) \cap X_i \neq \phi$ which implies  $n_A(x) \cap X_1 \neq \phi \lor n_A(x) \cap X_2 \neq \phi \lor \cdots \lor n_A(x) \cap X_r \neq \phi$ if and only if  $x \in \overline{R}^{\alpha}_{\delta}(X)$ , so  $\overline{R}^{\alpha}_{\delta}(X) = U$ ; (3) exists  $X_i \in \mathbf{X}, X_i = \phi$  implies for all  $x \in U$ , for all  $i \in \{1, 2, \dots, r\}, n_A(x) \notin X_i$ which implies  $x \notin \underline{R}^{\alpha}_{\delta}(X)$  so  $\underline{R}^{\alpha}_{\delta}(X) = \phi$ ; (4) for all  $X_i \in \mathbf{X}, X_i = \phi$  implies for all  $x \in U$ , for all  $i \in \{1, 2, \dots, r\}, n_A(x) \cap X_i = \phi$ which implies  $x \notin \overline{R}^{\alpha}_{\delta}(X)$  so  $\overline{R}^{\alpha}_{\delta}(X) = \phi$ ; (5) for all  $x \in \underline{R}^{\alpha}_{\delta}(X)$  implies for all  $i \in \{1, 2, \dots, r\}, n_A(x) \subseteq X_i$  which implies  $n_A(x) \subseteq X_1 \land n_A(x) \subseteq X_2 \land \dots \land n_A(x) \subseteq X_r$  so  $\underline{R}^{\alpha}_{\delta}(X) \subseteq X_i$ ; (6) for all  $x \in X_i$  implies for all  $i \in \{1, 2, \dots, r\}, n_A(x) \cap X_i \neq \phi$  which implies  $x \in \overline{R}^{\alpha}_{\delta}(X)$  so  $X_i \subseteq \overline{R}^{\alpha}_{\delta}(X)$ .  $\Box$ 

**Proposition 2.** Suppose U is a finite universe and (U, A) is an information system,  $\mathbf{Z} = \{Z_1, Z_2, \cdots, Z_r\}$  is a finite target set which satisfies for all  $Z_i \in \mathbf{Z}, Z_i \subseteq U$ .  $\mathbf{Z}$  is a *GCTS and the correlation coefficient is*  $\alpha$ *.X,Y*  $\subseteq$  *Z. For the approximations of NMTRS:* 

 $\begin{array}{l} (1) \ \underline{R}^{\alpha}_{\delta}(X \cap Y) \supseteq \underline{R}^{\alpha}_{\delta}(X) \cap \underline{R}^{\alpha}_{\delta}(Y); \\ (2) \ \overline{R}^{\alpha}_{\delta}(X \cap Y) \subseteq \overline{R}^{\alpha}_{\delta}(X) \cap \overline{R}^{\alpha}_{\delta}(Y); \end{array}$ 

- $\begin{array}{l} (3) \ \underline{R}^{\alpha}_{\delta}(X \cup Y) \subseteq \underline{R}^{\alpha}_{\delta}(X) \cup \underline{R}^{\alpha}_{\delta}(Y); \\ (4) \ \overline{R}^{\alpha}_{\delta}(X \cup Y) = \overline{R}^{\alpha}_{\delta}(X) \cup \overline{R}^{\alpha}_{\delta}(Y). \end{array}$

Proof.

(1) for all  $x \in \underline{R}^{\alpha}_{\delta}(X) \cap \underline{R}^{\alpha}_{\delta}(Y)$  implies for all  $X_i \in X(i \leq |X|), n_A(x) \subseteq X_i$ and for all  $Y_j \in Y(j \le |Y|)$ ,  $n_A(x) \subseteq Y_j$  which implies for all  $Z_k \in X \cap Y$  $(k \in \{1, 2, \cdots, r\}), n_A(x) \subseteq Z_k \text{ if and only if } x \in \underline{R}^{\alpha}_{\delta}(X \cap Y)$  $\therefore \underline{R}^{\alpha}_{\delta}(\mathbf{X} \cap \mathbf{Y}) \supseteq \underline{R}^{\alpha}_{\delta}(\mathbf{X}) \cap \underline{R}^{\alpha}_{\delta}(\mathbf{Y});$ (2) for all  $x \in \overline{R}^{\alpha}_{\delta}(X \cap Y)$ ,  $x \in \overline{R}^{\alpha}_{\delta}(X \cap Y)$  if and only if exists  $Z_k \in X \cap Y$  $(k \in \{1, 2, \dots, r\}), n_A(x) \cap Z_k \neq \phi$ , which implies exists  $X_i \in X(i \leq |X|), n_A(x) \cap X_i \neq \phi$ and exists  $Y_j \in \mathbf{Y}(j \leq |\mathbf{Y}|), n_A(x) \cap Y_j \neq \phi$ , which implies  $x \in \overline{R}^{\alpha}_{\delta}(\mathbf{X}) \cap \overline{R}^{\alpha}_{\delta}(\mathbf{Y})$  $\therefore \overline{R}^{\alpha}_{\delta}(\boldsymbol{X} \cap \boldsymbol{Y}) \subseteq \overline{R}^{\alpha}_{\delta}(\boldsymbol{X}) \cap \overline{R}^{\alpha}_{\delta}(\boldsymbol{Y});$ (3) for all  $x \in \underline{R}^{\alpha}_{\delta}(X \cup Y)$ ,  $x \in \underline{R}^{\alpha}_{\delta}(X \cup Y)$  if and only if exists  $Z_k \in X \cup Y$  $(k \in \{1, 2, \dots, r\}), n_A(x) \subseteq Z_k$ , which implies for all  $X_i \in \mathbf{X}(i \leq |\mathbf{X}|), n_A(x) \subseteq X_i$ or for all  $Y_j \in \mathbf{Y}(j \leq |\mathbf{Y}|), n_A(x) \subseteq Y_j$  which implies  $x \in \underline{R}^{\alpha}_{\delta}(\mathbf{X})$  or  $x \in \underline{R}^{\alpha}_{\delta}(\mathbf{Y})$ which implies  $x \in \underline{R}^{\alpha}_{\delta}(X) \cup \underline{R}^{\alpha}_{\delta}(Y)$  so  $\underline{R}^{\alpha}_{\delta}(X \cup Y) \subseteq \underline{R}^{\alpha}_{\delta}(X) \cup \underline{R}^{\alpha}_{\delta}(Y)$ ; (4) for all  $x \in \overline{R}^{\alpha}_{\delta}(X \cup Y)$ ,  $x \in \overline{R}^{\alpha}_{\delta}(X \cup Y)$  if and only if exists  $Z_k \in X \cup Y$  $(k \in \{1, 2, \dots, r\}), n_A(x) \cap Z_k \neq \phi$ , which implies exists  $X_i \in X(i \leq |X|), n_A(x) \cap X_i \neq \phi$ or exists  $Y_i \in Y(j \leq |Y|)$ ,  $n_A(x) \cap Y_j \neq \phi$ , which implies  $x \in \overline{R}^{\alpha}_{\delta}(X) \cup \overline{R}^{\alpha}_{\delta}(Y)$  $\therefore \overline{R}^{\alpha}_{\delta}(X \cup Y) = \overline{R}^{\alpha}_{\delta}(X) \cup \overline{R}^{\alpha}_{\delta}(Y). \ \Box$ 

**Proposition 3.** Suppose U is a finite universe and (U, A) is an information system,  $Z = \{Z_1, Z_2, \dots, Z_r\}$  is a finite target set which satisfies for all  $Z_i \in Z, Z_i \subseteq U$ . Z is a GCTS and the correlation coefficient is  $\alpha$ .  $X \subseteq Y \subseteq Z$ . For the approximations of NMTRS:

 $\begin{array}{l} (1) \ \underline{R}^{\alpha}_{\delta}(X) \supseteq \underline{R}^{\alpha}_{\delta}(Y); \\ (2) \ \overline{R}^{\alpha}_{\delta}(X) \subseteq \overline{R}^{\alpha}_{\delta}(Y). \end{array}$ 

Proof.

(1) for all  $x \in \underline{R}^{\alpha}_{\delta}(\mathbf{Y})$  if and only if for all  $Z_k \in \mathbf{Y}, k \in \{1, 2, \dots, r\}, n_A(x) \subseteq Z_k$ for all  $Z_k \in \mathbf{X}, k \in \{1, 2, \dots, r\}, n_A(x) \subseteq Z_k$ , implies  $x \in \underline{R}^{\alpha}_{\delta}(\mathbf{X})$ , so  $\underline{R}^{\alpha}_{\delta}(\mathbf{X}) \supseteq \underline{R}^{\alpha}_{\delta}(\mathbf{Y})$ ; (2) for all  $x \in \overline{R}^{\alpha}_{\delta}(\mathbf{X})$  if and only if for all  $Z_k \in \mathbf{X}, k \in \{1, 2, \dots, r\}, n_A(x) \cap Z_k \neq \phi$ , which implies exists  $Z_k \in \mathbf{Y}, k \in \{1, 2, \dots, r\}, n_A(x) \cap Z_k \neq \phi$ , which implies  $x \in \overline{R}^{\alpha}_{\delta}(\mathbf{Y})$ ,  $\therefore \overline{R}^{\alpha}_{\delta}(\mathbf{X}) \subseteq \overline{R}^{\alpha}_{\delta}(\mathbf{Y})$ .  $\Box$ 

# 2.3. Approximation Computation of NMTRS

In this section, we propose an approach for computing the approximations of NMTRS. We induce some corresponding results for approximation computation, then design an algorithm for calculating the approximations of NMTRS.

**Definition 6.** [36] Suppose U is a finite universe,  $P \subseteq U$ , the matrix representation of set P is defined as:

$$\stackrel{\rightarrow}{P} = (p_j)_{n \times 1}, j \in \{1, 2, \cdots, n\}, n = |U|, \text{ where } p_j = \begin{cases} 0 & x_j \in P\\ 1 & x_j \in P \end{cases}.$$

Example 2. Continues Example 1. By Definition 6 we have

$$\overrightarrow{n_A(x_1)} = (1,0,0,0,0)', \overrightarrow{n_A(x_2)} = (0,1,0,0,1)', \overrightarrow{n_A(x_3)} = (0,0,1,1,0)', \overrightarrow{n_A(x_3)} = (0,0,1,1,0)', \overrightarrow{n_A(x_5)} = (0,1,0,0,1)'.$$

where  $\dot{M}$  means the transpose of matrix  $\dot{M}$ .

**Lemma 1.** *Suppose*  $X, Y \subseteq U$ , we have

(1)  $X \cap Y \neq \phi$  if and only if  $\vec{X}' \cdot \vec{Y} > 0$ ;

(2) 
$$X \subseteq Y$$
 if and only if  $X \cdot \sim Y = 0$ 

where  $\sim Y$  is a complement of Y, where  $\overrightarrow{X}'$  means the transpose of a matrix  $\overrightarrow{X}$ , and '.' is the matrix quantity product.

There is a useful property gained by computing the approximations of NMTRS using Lemma 1.

**Theorem 1.** Suppose U is a finite universe and (U, A) is an information system, **X** is a GCTS, and we have

(1) 
$$x \in \underline{R}^{\alpha}_{\delta}(\mathbf{X})$$
 if and only if for all  $i \in \{1, 2, \dots, r\}, \overline{n_A(x)}, \overline{\langle X_i \rangle} = 0;$   
(2)  $x \in \overline{R}^{\alpha}_{\delta}(\mathbf{X})$  if and only if exists  $i \in \{1, 2, \dots, r\}, \overline{n_A(x)}, \overline{X_i} > 0.$ 

**Proof.** From Lemma 1,

(1) for all  $i \in \{1, 2, \dots, r\}$ ,  $\overrightarrow{n_A(x)}' \cdot \overrightarrow{\sim X_i} = 0$  which implies for all  $i \in \{1, 2, \dots, r\}$ ,  $n_A(x) \subseteq X_i$ , if and only if  $x \in \underline{R}^{\alpha}_{\delta}(\mathbf{X})$ ; (2) exists  $i \in \{1, 2, \dots, r\}$ ,  $\overrightarrow{n_A(x)}' \cdot \overrightarrow{X_i} > 0$ , which implies exits  $i \in \{1, 2, \dots, r\}$ ,  $n_A(x) \cap X_i \neq \phi$ , if and only if  $x \in \overline{R}^{\alpha}_{\delta}(\mathbf{X})$ .  $\Box$ 

The matrix representation of Theorem 1 is as the following definition.

**Definition 7.** Suppose U is a finite universe and (U, A) is an information system, X is a GCTS, and the lower approximation matrix of NMTRS can be defined as:

$$\underline{H}(\mathbf{X}) = \left(\underline{h}_{ij}\right)_{r \times n'} \text{ where } \underline{h}_{ij} = \begin{cases} 1 & \overrightarrow{n_A(x_j)'} \cdot \overrightarrow{\sim X_i} = 0, i \in \{1, 2, \cdots, r\}, j \in \{1, 2, \cdots, n\} \\ 0 & \text{otherwise} \end{cases}$$

The upper approximation matrix of NMTRS can be defined as:

$$\overline{H}(\mathbf{X}) = \left(\overline{h}_{ij}\right)_{r \times n'} \text{ where } \overline{h}_{ij} = \begin{cases} 1 & \overrightarrow{n_A(x_j)'} \cdot \overrightarrow{X}_i > 0, i \in \{1, 2, \cdots, r\}, j \in \{1, 2, \cdots, n\} \\ 0 & \text{otherwise} \end{cases}$$

Example 3. Continuation of Example 2. By Definition 6 we have

$$\vec{X}_1 = (0, 1, 1, 0, 1)', \ \vec{\nabla X_1} = (1, 0, 0, 1, 0)', \ \vec{X}_2 = (1, 1, 0, 0, 1)', \ \vec{\nabla X_2} = (0, 0, 1, 1, 0)'.$$

By Definition 7 we have

$$\vec{n_A(x_1)}' \xrightarrow{\sim} \vec{X_1} = (1,0,0,0,0) \cdot (1,0,0,1,0)' \neq 0,$$
  

$$\vec{n_A(x_1)}' \xrightarrow{\sim} \vec{X_2} = (1,0,0,0,0) \cdot (0,0,1,1,0)' = 0;$$
  

$$\vec{n_A(x_2)}' \xrightarrow{\sim} \vec{X_1} = 0, \vec{n_A(x_2)}' \xrightarrow{\sim} \vec{X_2} = 0; \vec{n_A(x_3)}' \xrightarrow{\sim} \vec{X_1} \neq 0, \vec{n_A(x_3)}' \xrightarrow{\sim} \vec{X_2} \neq 0;$$
  

$$\vec{n_A(x_4)}' \xrightarrow{\sim} \vec{X_1} \neq 0, \vec{n_A(x_4)}' \xrightarrow{\sim} \vec{X_2} \neq 0; \vec{n_A(x_5)}' \xrightarrow{\sim} \vec{X_1} = 0, \vec{n_A(x_5)}' \xrightarrow{\sim} \vec{X_2} = 0;$$
  

$$\vec{n_A(x_4)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

then 
$$\underline{H}(\mathbf{X}) = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$
.

Example 4. Continuation of Example 3. By Definition 7 we have

$$\overrightarrow{n_A(x_1)}' \cdot \overrightarrow{X_1} = (1,0,0,0,0) \cdot (0,1,1,0,1)' = 0, \overrightarrow{n_A(x_1)}' \cdot \overrightarrow{X_2} = (1,0,0,0,0) \cdot (1,1,0,0,1)' > 0; \overrightarrow{n_A(x_2)}' \cdot \overrightarrow{X_1} > 0, \overrightarrow{n_A(x_2)}' \cdot \overrightarrow{X_2} > 0; \overrightarrow{n_A(x_3)}' \cdot \overrightarrow{X_1} > 0, \overrightarrow{n_A(x_3)}' \cdot \overrightarrow{X_2} = 0; \overrightarrow{n_A(x_4)}' \cdot \overrightarrow{X_1} = 0, \overrightarrow{n_A(x_4)}' \cdot \overrightarrow{X_2} = 0; \overrightarrow{n_A(x_5)}' \cdot \overrightarrow{X_1} > 0, \overrightarrow{n_A(x_5)}' \cdot \overrightarrow{X_2} > 0;$$

then  $\overline{H}(\mathbf{X}) = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$ .

Then, we can easily obtain a theorem for computing approximations in NMTRS. **Theorem 2.** Suppose U is a finite universe and (U, A) is an information system, X is a GCTS,  $i \in \{1, 2, \dots, r\}, j \in \{1, 2, \dots, n\}$ , we have

(1) 
$$x_j \in \underline{R}^{\alpha}_{\delta}(\mathbf{X})$$
 if and only if  $\wedge_{i=1}^r \underline{h}_{ij} = 1$ ;  
(2)  $x_j \in \overline{R}^{\alpha}_{\delta}(\mathbf{X})$  if and only if  $\vee_{i=1}^r \overline{h}_{ij} = 1$ .

Proof. From Lemma 1,

(1) 
$$x_j \in \underline{R}^{\alpha}_{\delta}(\mathbf{X})$$
 if and only if for all  $i \in \{1, 2, \dots, r\}, \overline{n_A(x_j)}, \overline{\langle X_i \rangle} = 0$   
if and only if for all  $i \in \{1, 2, \dots, r\}, \underline{h}_{ij} = 1, \text{so } \wedge_{i=1}^r \underline{h}_{ij} = 1;$   
(2)  $x_j \in \overline{R}^{\alpha}_{\delta}(\mathbf{X})$  if and only if exists  $i \in \{1, 2, \dots, r\}, \overline{n_A(x_j)}, \overline{X}_i > 0,$   
if and only if exists  $i \in \{1, 2, \dots, r\}, \overline{h}_{ij} = 1, \text{so } \vee_{i=1}^r \overline{h}_{ij} = 1.$ 

**Example 5.** Continuation of Example 4. By Theorem 2 we have

$$\overrightarrow{R_{\delta}^{\alpha}(\mathbf{X})} = (0, 1, 0, 0, 1)' \land (1, 1, 0, 0, 1)' = (0, 1, 0, 0, 1)',$$
  

$$\overrightarrow{R_{\delta}^{\alpha}(\mathbf{X})} = (0, 1, 1, 0, 1)' \lor (1, 1, 0, 0, 1)' = (1, 1, 1, 0, 1)',$$
  

$$\therefore \underline{R_{\delta}^{\alpha}(\mathbf{X})} = \{x_2, x_5\}, \overline{R_{\delta}^{\alpha}(\mathbf{X})} = \{x_1, x_2, x_3, x_5\}.$$

# 3. Attribute Reduction Based on NMTRS

In this section, we propose an attribute reduction method based on NMTRS. We derive the attribute significance measure, then design an algorithm for calculating the reduction.

### 3.1. Attribute Significance Measure Based on NMTRS

Suppose  $U = \{x_1, x_2, \dots, x_n\}$  is the instance space and  $L = \{l_1, l_2, \dots, l_r\}$  is the label space.  $T = \{(x_i, y_i) | i = 1, 2, \dots, n\}$  means that label vector  $y_i$  is associated with instance  $x_i$ . (U, A, L) is the tuple denoting the multi-label information system.

**Definition 8.** Suppose U is a finite universe and (U, A, L) is an information system,  $X = \{X_1, X_2, \dots, X_m\}$  is a GCTS and the correlation coefficient is  $\alpha$ . If we regard  $l_i$   $(i = 1, 2, \dots, r)$  as the matrix representation of a certain decision set, and  $\phi(L) = \{X^1, X^2, \dots\}$ is the GCTS group on L, for any  $B \subseteq A$  then the positive region of NMTRS is defined as:

$$Pos_B(\phi(L)) = \{ \cup R_B^{\alpha}_{\delta}(X) | X \in \phi(L) \},\$$

then, the dependence of NMTRS on attribute set B can be defined as:

$$\gamma_B(\phi(L)) = \frac{|Pos(\phi(L))|}{|U|},$$

then, for all  $c \in A - B$ , the conditional attribute significance of c on attribute set B can be calculated as:

$$sig_{\gamma}(c, B, L) = \gamma_{B \cup \{c\}}(\phi(L)) - \gamma_B(\phi(L)).$$

It can be easily obtained that  $sig_{\gamma}(c, B, L) \ge 0$ , and if  $sig_{\gamma}(c, B, L) = 0$ , we can say that c is an unnecessary attribute.

Based on Theorem 2, we propose matrix-based Algorithm 1 for computing the approximations of a particular target concept group. The total time complexity of Algorithm 1 is  $\Theta(rn^2)$ . Steps 2–15 are to calculate  $\overline{H}$  and  $\underline{H}$  with time complexity  $\Theta(rn^2)$ .

Algorithm 1: Computing approximations of NMTRS

**Input:**  $(U, A), \mathbf{X} = \{X_1, X_2, \cdots, X_r\}, \overrightarrow{n_A(x_j)}, \text{ for all } x \in U$ **Output:**  $\overline{R}^{\alpha}_{\delta}(X), \underline{R}^{\alpha}_{\delta}(X)$ 1:  $n \leftarrow |U|, r \leftarrow |X|$ 2: for  $i = 1 \rightarrow r$ if  $\overrightarrow{n_A(x_j)}' \xrightarrow{i} \sim X_i = 0$  then  $\underline{h_{ij}} = 1$ 3: for  $j = 1 \rightarrow n$ 4: 5: 6: else 7:  $\underline{h}_{ij} = 0$ end if 8: if  $\overrightarrow{n_A(x_j)} \cdot \overrightarrow{X_i} > 0$  then 9:  $\overline{h}_{ij} = 1$ 10: 11: else  $\overline{h}_{ij} = 0$ 12: 13: end if end for  $14 \cdot$ 15: end for 16:  $\overline{\overline{R}_{\delta}^{\alpha}(\mathbf{X})}' = \left(\wedge_{i=1}^{r}\underline{h}_{i*}\right)' = \left(\wedge_{i=1}^{r}\underline{h}_{i1}, \wedge_{i=1}^{r}\underline{h}_{i2}, \cdots, \wedge_{i=1}^{r}\underline{h}_{in}\right)'$ 17:  $\underline{\overline{R}_{\delta}^{\alpha}(\mathbf{X})}' = \left(\vee_{i=1}^{r}\overline{h}_{i*}\right)' = \left(\vee_{i=1}^{r}\overline{h}_{i1}, \vee_{i=1}^{r}\overline{h}_{i2}, \cdots, \vee_{i=1}^{r}\overline{h}_{in}\right)'$ 18: Return  $\overline{R}^{\alpha}_{\delta}(X), \underline{R}^{\alpha}_{\delta}(X)$ 

### 3.2. Multi-Granulation Discrimination Analysis Based on NMTRS

Based on the NMTRS model, some discussions about multi-granulation discrimination are conducted in this section. Several propositions can be easily obtained.

**Proposition 4.** Suppose U is a finite universe and (U, A, L) is an information system,  $X = \{X_1, X_2, \dots, X_m\}$  is a GCTS and the correlation coefficient is  $\alpha$ . d is the Euclidean distance,  $B_1 \subseteq B_2 \subseteq A$ , and the following properties can be easily obtained:

(1) for all  $x \in U$ ,  $n_{B_1}(x) \subseteq n_{B_2}(x)$ ; (2) for all  $\mathbf{X} \in \phi(L)$ ,  $R_{B_1}^{\alpha}(\mathbf{X}) \subseteq R_{B_2}^{\alpha}(\mathbf{X})$ ; (3)  $Pos_{B_1}(\phi(L)) \subseteq Pos_{B_2}^{\delta}(\phi(L))$ ; (4)  $\gamma_{B_1}(\phi(L)) \leq \gamma_{B_2}(\phi(L))$ .

**Proposition 5.** Suppose U is a finite universe and (U, A, L) is an information system,  $X = \{X_1, X_2, \dots, X_m\}$  is a GCTS and the correlation coefficient is  $\alpha$ . d is the Euclidean distance, and for any  $B_1 \subseteq B_2 \subseteq A$ , if  $x \in Pos_{B_1}(\phi(L))$ , then  $x \in Pos_{B_2}(\phi(L))$ . **Proposition 6.** Suppose U is a finite universe and (U, A, L) is an information system,  $X = \{X_1, X_2, \dots, X_m\}$  is a GCTS and the correlation coefficient is  $\alpha$ . d is the Euclidean distance, and for any  $0 \le \delta_2 \le \delta_1$ ,  $B \subseteq A$ , the following properties can be easily obtained:

(1) for all  $x \in U$ ,  $n_B^{\delta_2}(x) \subseteq n_B^{\delta_1}(x)$ ; (2) for all  $\mathbf{X} \in \phi(L)$ ,  $R_B^{\alpha}(\mathbf{X}) \subseteq R_B^{\alpha}(\mathbf{X})$  and  $\overline{R_B}^{\alpha}_{\delta_2}(\mathbf{X}) \subseteq \overline{R_B}^{\alpha}_{\delta_1}(\mathbf{X})$ ; (3)  $Pos_{\delta_1}(\phi(L)) \subseteq Pos_{\delta_2}(\phi(L))$ ; (4)  $\gamma_{\delta_1}(\phi(L)) \leq \gamma_{\delta_2}(\phi(L))$ .

**Proposition 7.** Suppose U is a finite universe and (U, A, L) is an information system,  $L' \subseteq L$ . *d* is the Euclidean distance, and for any  $0 \le \delta_2 \le \delta_1$ , the following properties can be easily obtained:

(1)  $Pos_{\delta_1}(\phi(L)) \subseteq Pos_{\delta_2}(\phi(L'));$ (2)  $\gamma_{\delta_1}(\phi(L)) \leq \gamma_{\delta_2}(\phi(L')).$ 

#### 3.3. Attribute Reduction Algorithm Based on NMTRS

Based on the NMTRS model and its attribute significance measure, we derive the corresponding attribute reduction algorithm, which is named Algorithm 2.

Algorithm 2: Attribute reduction algorithm based on NMTRS

```
Input: (U, A, L), \alpha, \delta, \phi(L)
Output: Reduction of the attribute set A
1: Reduction \leftarrow \emptyset
2: while A - Reduction \neq \emptyset
3:
          for all b \in A - Reduction
4:
                Calculate sig_{\gamma}(b, Reduction, L) with Algorithm 1
5:
          end for
6: if \max_{b \in A-Reduction} \{ sig_{\gamma}(b, Reduction, L) \} \le 0 then
7:
         break
8: end if
          Reduction = Reduction \cup \underset{b \in A-Reduction}{\operatorname{argmax}} \{ sig_{\gamma}(b, Reduction, L) \}
9:
10: Return Reduction
```

The time complexity of Algorithm 2 is  $\Theta(rn^2)$ . The time complexity of calculating the neighborhood class of all instances is  $\Theta(n \times \log n)$ , and the time complexity of calculating the attribute significance of attributes is  $\Theta(rn^2)$ .

# 4. Experimental Evaluations

In this section, several experiments are conducted to evaluate the effectiveness and the efficiency of the algorithms we proposed, namely Algorithm 2 (ours), Algorithm 3 (MLNB), Algorithm 4 (Laplacian Score) [37], Algorithm 5 (RelieF) [38], Algorithm 6 MDDM proj) and Algorithm 7 (MDDMspc) [24]. Since we can only determine the attribute order from MLNB and Laplacian Score, we make the number of the chosen attributes the same as our algorithm. Six datasets were chosen from public repositories. The details of these datasets are listed in Table 2. All the experiments were carried out on a personal computer with 64-bit Windows 10, Inter(R) Core (TM) i7 1065G7 CPU@3.50GHz and 16GB memory, developed by Wenbin Zheng, in Zhangzhou, China. The programming language was MATLAB r2020a.

No.	Datasets	Samples	Attributes	Labels
1	Bird Song	4998	38	13
2	CAL-500	502	68	174
3	Emotions	593	72	6
4	Flags	194	14	12
5	FGNET	1002	262	78
6	Water Quality Nom	1060	16	14

Table 2. Details of datasets.

## 4.1. Comparison of Performance Measures Using ML-kNN

In this subsection, we use ML-kNN as the learning method and select features using the four attribute reduction approaches, conducting ten-fold cross-validation on the temporary dataset ten times with the selected feature set.

### 4.1.1. Experimental Settings

For comparing the performance of our approach with the comparison algorithms, we apply all of the attribute reduction algorithms on the six datasets using their recommended parameter configurations in [39]. For our method, we set  $\alpha = 0.6$  and  $\delta = 0.5$ . The experimental results are shown in Tables 3–8.

Table 3. Experimental results on Bird Song.

Algorithms	Hamming Loss (↓)	Ranking Loss (↓)	One-Error (↓)	Coverage (↓)	Average Precision (↑)
Ours	1	_	0.60024	1.3229	0.61599
MLNB	—	-	—	-	-
Laplacian Score	1	—	0.60024	1.3229	0.61599
ReliefF	1	-	0.60024	1.3229	0.61599
MDDMproj	0.94641	—	0.22601	0.41394	0.86440
MDDMspc	0.94641	_	0.22601	0.41394	0.86440

 Table 4. Experimental results on CAL-500.

Algorithms	Hamming Loss (↓)	Ranking Loss (↓)	One-Error (↓)	Coverage ( $\downarrow$ )	Average Precision (↑)
Ours	0.96755	0.18771	0.13001	132.03	0.48148
MLNB	0.96806	0.18763	0.12386	131.26	0.48137
Laplacian Score	0.96880	0.18701	0.11717	131.22	0.48484
RelieF	0.96870	0.18593	0.1208	131	0.48684
MDDMproj MDDMspc	0.96755 0.96806	0.18771 0.18763	0.13001 0.12386	132.03 131.26	$0.48148 \\ 0.48137$

Table 5. Experimental results on Emotions.

Algorithms	Hamming Loss (↓)	Ranking Loss (↓)	One-Error (↓)	Coverage (↓)	Average Precision (†)
Ours	0.95468	0.51255	2.7671	0.62310	0.95468
MLNB	—	—	—	_	—
Laplacian Score	0.95033	0.56044	2.9738	0.59137	0.95033
RelieF	0.89657	0.40958	2.4069	0.69350	0.89657
MDDMproj MDDMspc	0.89657 0.88908	0.40958 0.42924	2.4069 2.3618	0.69350 0.68983	0.89657 0.88908

Algorithms	Hamming Loss (↓)	Ranking Loss (↓)	One-Error (↓)	Coverage (↓)	Average Precision (↑)
Ours	0.81977	0.24048	6.1901	0.75653	0.81977
MLNB	—	—	—	—	—
Laplacian Score	0.80916	0.23298	5.8892	0.76737	0.80916
RelieF	0.81433	0.23269	6.2364	0.75832	0.81433
MDDMproj MDDMspc	$0.81416 \\ 0.81416$	$0.21904 \\ 0.21904$	6.1963 6.1963	0.76229 0.76229	$0.81416 \\ 0.81416$

Table 6. Experimental results on Flags.

Table 7. Experimental results on FGNET.

Algorithms	Hamming Loss (↓)	Ranking Loss (↓)	One-Error (↓)	Coverage ( $\downarrow$ )	Average Precision (↑)
Ours	1	_	0.93427	21.477	0.15097
MLNB	_	_	_	_	—
Laplacian Score	1	—	0.93427	21.477	0.15097
RelieF	1	_	0.93427	21.477	0.15097
MDDMproj	1	_	0.95291	21.475	0.13413
MDDMspc	1	—	0.95291	21.475	0.13413

Table 8. Experimental results on Water Quality Nom.

Algorithms	Hamming Loss (↓)	Ranking Loss (↓)	One-Error (↓)	Coverage ( $\downarrow$ )	Average Precision (↑)
Ours	0.88209	_	0.33826	9.3556	_
MLNB	_	_	_	_	_
Laplacian Score	0.8825	_	0.35662	9.2995	_
RelieF	0.85437	_	0.32326	9.2675	_
MDDMproj	0.85437	_	0.32326	9.2675	_
MDDMspc	0.88296	_	0.36373	9.5964	_

4.1.2. Discussions of the Experimental Results

From Tables 3–8, we can easily observe that our algorithm has considerable performance on each dataset, while none of the algorithms have obvious superiority on all datasets. Sometimes our algorithm can reach the best performance; it is in the middle class generally. The experiments show that our algorithm is valid so it does provide a new approach for multi-label feature selection.

For the dataset Bird Song, MDDM algorithms perform better than other algorithms; the performance of our algorithm is equal to MLNB, RelieF, and Laplacian Score.

For the dataset CAL-500, our algorithm performs equally to MDDMproj, better than all other algorithms with the performance measure Hamming loss.

For the dataset Emotions, our algorithm is the worst one with Hamming loss, ranking loss, and Average Precision, but performs better than RelieF, MDDMproj, and MDDMspec with the performance measure Coverage.

For the dataset FGNET, MDDM algorithms perform better, while our algorithm's performance is equal to other algorithms.

For the dataset Water Quality Nom, our algorithm performs better than Laplacian Score and MDDMspc with the performance measure Hamming loss, performs better than Laplacian Score and MDDMspc with One-Error, and performs better than MDDMspc with Coverage.

# 4.2. Effectiveness of Parameters $\alpha$ and $\delta$

In this subsection, we conduct several experiments to analyze the effectiveness of  $\alpha$  and  $\delta$ .

# 4.2.1. Experimental Settings

For analyzing the effectiveness of parameters, we conduct several experiments. We gradually increase the value of parameter  $\alpha$  from 0 to 1 with a step of 0.05 and  $\delta$  from 0 to  $10^{20}$  with a step of  $10^{1}$ . The experimental result is shown in Figure 1.



**Figure 1.** Effectiveness of parameter  $\alpha$  and  $\delta$  on dataset Emotions.

### 4.2.2. Discussions of the Experimental Results

In Figure 1 we can easily observe that with the increase in  $\alpha$ , the performance of our proposed algorithm decreases when  $\alpha$  is close to 1. Parameter  $\delta$  does not have a significant effect on performance. It provides the recommended parameter setting for our experiments.

#### 4.3. Effectiveness of Noise

In this subsection, we conduct several experiments to analyze the effectiveness of noise.

#### 4.3.1. Experimental Settings

To analyze the effectiveness of noise, we conduct several experiments. We gradually increase the percentage of noise in the dataset Flags from 0 to 1 with a step of 0.05. The experimental result is shown in Figure 2.



Figure 2. Robustness while adding noise into the dataset Flags.

#### 4.3.2. Discussions of the Experimental Results

In Figure 1 we can easily observe that all the algorithms display similar performance with the increasing noise percentage. Compared with another algorithm, our proposed algorithm is not worse than them. In other words, for robustness, the performance of our algorithm is equivalent to comparison algorithms.

# 5. Conclusions

Attribute reduction in multi-label learning has been a hot topic in recent years. We propose a novel rough set model for multi-label learning and investigate the properties of the proposed model. With the proposed model, a novel feature selection method for multi-label learning is proposed, a novel attribute reduction approach is provided, and its validation is proved by experiments.

In the big data era, calculating the reduction of an attribute set presents great challenges for its high time complexity and dynamic attribute reduction. Algorithms can solve the problem by using previously calculated results. In future work, we will carry on promoting the proposed approach into a dynamic multi-label attribute reduction approach.

**Author Contributions:** Conceptualization, W.Z. and S.L. and Y.L.; methodology, W.Z. and J.L.; software, W.Z.; validation, W.Z., S.L. and Y.L.; formal analysis, W.Z. and J.L.; investigation, W.Z. and S.L.; writing—review and editing, W.Z.; supervision, J.L. and S.L.; project administration, W.Z. and S.L.; funding acquisition, S.L. and J.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is supported in part by the National Natural Science Foundation of China under Grant Nos. 11871259, 61379021 and 12101289, the Natural Science Foundation of Fujian Province under Nos. 2022J01912 and 2022J01306, the Institute of Meteorological Big Data-Digital Fujian, Fujian Key Laboratory of Data Science and Statistics, and Fujian Key Laboratory of Granular Computing and Applications.

**Data Availability Statement:** The data that support the findings of this study are openly available in UC Irvine Machine Learning Repository at https://archive.ics.uci.edu/ml/index.php, http://mulan.sourceforge.net/datasets-mlc.html and https://www.uco.es/kdis/mllresources/, access on 16 April 2022.

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

#### References

- 1. Pawlak, Z. Rough set. Int. J. Comput. Inf. Sci. 1982, 11, 341–356. [CrossRef]
- 2. Ziarko, W. Probabilistic approach to rough sets. Int. J. Approx. Reason. 2008, 49, 272–284. [CrossRef]
- Bazan, J.G.; Nguyen, H.S.; Nguyen, S.H.; Synak, P.; Wróblewski, J. Rough set algorithms in classification problem. In Rough Set Methods and Applications Physical; Springer: Heidelberg, Germany, 2000.
- 4. Lingras, P. Unsupervised Rough Set Classification Using Gas. J. Intell. Inf. Syst. 2001, 16, 215–228. [CrossRef]
- Miao, D.; Duan, Q.; Zhang, H.; Jiao, N. Rough set based hybrid algorithm for text classification. *Expert Syst. Appl.* 2009, 36, 9168–9174. [CrossRef]
- Sharma, H.K.; Majumder, S.; Biswas, A.; Prentkovskis, O.; Kar, S.; Skačkauskas, P. A Study on Decision-Making of the Indian Railways Reservation System during COVID-19. *J. Adv. Transp.* 2022, 2022, 7685375. Available online: https://www.hindawi. com/journals/jat/2022/7685375/ (accessed on 7 July 2022). [CrossRef]
- Lingras, P. Rough set clustering for web mining. In Proceedings of the 2002 IEEE World Congress on Computational Intelligence, 2002 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291), Honolulu, HI, USA, 12–17 May 2002.
- Lingras, P.; Peters, G. Applying rough set concepts to clustering. In Rough Sets: Selected Methods and Applications in Management and Engineering; Springer: London, UK, 2012; pp. 23–37.
- Parmar, D.; Wu, T.; Blackhurst, J. MMR: An algorithm for clustering categorical data using Rough Set Theory. *Data Knowl. Eng.* 2007, 63, 879–893. [CrossRef]
- 10. Vidhya, K.A.; Geetha, T.V. Rough set theory for document clustering: A review. J. Intell. Fuzzy Syst. 2017, 32, 2165–2185. [CrossRef]
- Hedar, A.R.; Ibrahim, A.M.M.; Abdel-Hakim, A.E.; Sewisy, A.A. Modulated clustering using integrated rough sets and scatter search attribute reduction. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, Kyoto, Japan, 15–19 July 2018; pp. 1394–1401.
- 12. Xia, S.; Zhang, H.; Li, W.; Wang, G.; Giem, E.; Chen, Z. GBNRS: A Novel Rough Set Algorithm for Fast Adaptive Attribute Reduction in Classification. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 1231–1242. [CrossRef]
- 13. Qian, Y.; Liang, J.; Yao, Y.; Dang, C. MGRS: A multi-granulation rough set. Inf. Sci. 2010, 180, 949–970. [CrossRef]
- 14. Yao, Y.; Yao, B. Covering based rough set approximations. Inf. Sci. 2012, 200, 91–107. [CrossRef]
- Kumar, S.U.; Inbarani, H.H. A Novel Neighborhood Rough Set Based Classification Approach for Medical Diagnosis. *Procedia* Comput. Sci. 2015, 47, 351–359. [CrossRef]

- 16. Zhang, J.; Li, T.; Ruan, D.; Liu, D. Neighborhood rough sets for dynamic data mining. *Int. J. Intell. Syst.* **2012**, 27, 317–342. [CrossRef]
- 17. Hu, Q.; Yu, D.; Liu, J.; Wu, C. Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* 2008, 178, 3577–3594. [CrossRef]
- 18. Yong, L.; Wenliang, H.; Yunliang, J.; Zeng, Z. Quick attribute reduce algorithm for neighborhood rough set model. *Inf. Sci.* 2014, 271, 65–81. [CrossRef]
- 19. Chen, H.; Li, T.; Cai, Y.; Luo, C.; Fujita, H. Parallel attribute reduction in dominance-based neighborhood rough set. *Inf. Sci.* 2016, 373, 351–368. [CrossRef]
- Zhou, P.; Hu, X.; Li, P.; Wu, X. Online streaming feature selection using adapted Neighborhood Rough Set. Inf. Sci. 2018, 481, 258–279. [CrossRef]
- 21. Sun, L.; Ji, S.; Ye, J. Multi-Label Dimensionality Reduction; CRC Press: Boca Raton, FL, USA, 2016. [CrossRef]
- 22. Fisher, R.A. The use of multiple measurements in taxonomic problems. Ann. Eugen. 1936, 7, 179–188. [CrossRef]
- 23. Wold, H. Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis;* Academic Press: Cambridge, MA, USA, 1966; pp. 391–420.
- Zhang, Y.; Zhou, Z.H. Multilabel dimensionality reduction via dependence maximization. ACM Trans. Knowl. Discov. Data 2010, 4, 1–21. [CrossRef]
- Zhang, M.-L.; Peña, J.M.; Robles, V. Feature selection for multi-label naïve Bayes classification. *Inf. Sci.* 2009, 179, 3218–3229. [CrossRef]
- Lin, Y.; Hu, Q.; Liu, J.; Chen, J.; Duan, J. Multi-label feature selection based on neighborhood mutual information. *Appl. Soft Comput.* 2016, 38, 244–256. [CrossRef]
- 27. Liu, J.; Lin, Y.; Li, Y.; Weng, W.; Wu, S. Online multi-label streaming feature selection based on neighborhood rough set. *Pattern Recognit.* **2018**, *84*, 273–287. [CrossRef]
- 28. Deng, Z.; Zheng, Z.; Deng, D.; Wang, T.; He, Y.; Zhang, D. Feature Selection for Multi-Label Learning Based on F-Neighborhood Rough Sets. *IEEE Access* 2020, *8*, 39678–39688. [CrossRef]
- 29. Al-Shami, T.M.; Ciucci, D. Subset neighborhood rough sets. Knowl. Based Syst. 2021, 237, 107868. [CrossRef]
- Chen, Y.; Xue, Y.; Ma, Y.; Xu, F. Measures of uncertainty for neighborhood rough sets. *Knowl. Based Syst.* 2017, 120, 226–235. [CrossRef]
- 31. Wang, C.; Shi, Y.; Fan, X.; Shao, M. Attribute reduction based on k-nearest neighborhood rough sets. *Int. J. Approx. Reason.* 2018, 106, 18–31. [CrossRef]
- 32. Li, Y.; Lin, Y.; Liu, J.; Weng, W.; Shi, Z.; Wu, S. Feature selection for multi-label learning based on kernelized fuzzy rough sets. *Neurocomputing* **2018**, *318*, 271–286. [CrossRef]
- 33. Xu, J.; Shen, K.; Sun, L. Multi-label feature selection based on fuzzy neighborhood rough sets. *Complex Intell. Syst.* 2022, *8*, 2105–2129. [CrossRef]
- 34. Wang, Q.; Qian, Y.; Liang, X.; Guo, Q.; Liang, J. Local neighborhood rough set. Knowl. Based Syst. 2018, 153, 53–64. [CrossRef]
- Lin, G.; Qian, Y.; Li, J. NMGRS: Neighborhood-based multi-granulation rough sets. Int. J. Approx. Reason. 2012, 53, 1080–1093. [CrossRef]
- 36. Ziarko, W. Variable precision rough set model. J. Comput. Syst. Sci. 1993, 46, 39–59. [CrossRef]
- He, X.; Cai, D.; Niyogi, P. Laplacian score for feature selection. *Adv. Neural Inf. Process. Syst.* 2005, 18, 1–8. Available online: https://proceedings.neurips.cc/paper/2005/file/b5b03f06271f8917685d14cea7c6c50a-Paper.pdf (accessed on 7 July 2022).
- Spolaôr, N.; Cherman, E.A.; Monard, M.C.; Lee, H.D. Relief for multi-label feature selection. In Proceedings of the 2013 Brazilian Conference on Intelligent Systems, Fortaleza, Brazil, 20–24 October 2013; IEEE: Manhattan, NY, USA, 2013; pp. 6–11.
- 39. Liu, G.L. Axiomatic systems for rough sets and fuzzy rough sets. Int. J. Approx. Reason. 2008, 48, 857–867. [CrossRef]