

## Article

# TFD-IIS-CRMCB: Telecom Fraud Detection for Incomplete Information Systems Based on Correlated Relation and Maximal Consistent Block

Ran Li <sup>1</sup>, Hongchang Chen <sup>2,\*</sup>, Shuxin Liu <sup>2,\*</sup>, Kai Wang <sup>2</sup>, Biao Wang <sup>1</sup> and Xinxin Hu <sup>1</sup>

<sup>1</sup> Institute of Information Technology, PLA Strategic Support Force Information Engineering University, Zhengzhou 450002, China

<sup>2</sup> National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China

\* Correspondence: ndscchc@139.com (H.C.); liushxinll@126.com (S.L.)

**Abstract:** Telecom fraud detection is of great significance in online social networks. Yet the massive, redundant, incomplete, and uncertain network information makes it a challenging task to handle. Hence, this paper mainly uses the correlation of attributes by entropy function to optimize the data quality and then solves the problem of telecommunication fraud detection with incomplete information. First, to filter out redundancy and noise, we propose an attribute reduction algorithm based on max-correlation and max-independence rate (MCIR) to improve data quality. Then, we design a rough-gain anomaly detection algorithm (MCIR-RGAD) using the idea of maximal consistent blocks to deal with missing incomplete data. Finally, the experimental results on authentic telecommunication fraud data and UCI data show that the MCIR-RGAD algorithm provides an effective solution for reducing the computation time, improving the data quality, and processing incomplete data.

**Keywords:** telecom fraud detection; attribute reduction; incomplete information system; maximal consistent block; MCIR-RGAD



**Citation:** Li, R.; Chen, H.; Liu, S.; Wang, K.; Wang, B.; Hu, X. TFD-IIS-CRMCB: Telecom Fraud Detection for Incomplete Information Systems Based on Correlated Relation and Maximal Consistent Block. *Entropy* **2023**, *25*, 112. <https://doi.org/10.3390/e25010112>

Academic Editors: Carlos Alberto De Bragança Pereira, Adriano Polpo, Agatha Rodrigues and Débora Corrêa

Received: 1 November 2022

Revised: 24 December 2022

Accepted: 3 January 2023

Published: 5 January 2023



**Copyright:** © 2023 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

The digital age has dramatically facilitated many aspects of our lives, whereas cybersecurity issues threaten the positive effects of technology. Since unsafe information and illegitimate users blend so well with regular information and users that they can hardly be distinguished, cybersecurity threats [1] especially online fraud, telecommunications fraud [2], online social network fraud [3], credit card fraud [4], bank fraud [5], and fraudulent credit applications [6], have become a knotty governance problem.

Fraud detection [7] is a kind of anomaly detection and is usually tackled as a classification problem by screening abnormal items out with traditional machine learning methods [8,9] or deep learning ones [10–12]. Compared with the traditional machine learning model, the deep learning model has the problems of poor interpretability and no direction for parameter adjustment, and its calculation time increases with the complexity index of the model. Traditional machine learning is still widely studied and applied because of its strong interpretability and fast computing speed. The traditional outlier detection methods are mainly based on distribution-based [13], distance-based [14], density-based [15], and clustering-based [16] perspectives. However, traditional approaches to anomaly detection rely heavily on the relevance of features to the classification task. When the feature space is large, the presence of invalid, irrelevant, redundant, or noisy attributes in the data may inevitably affect the performance of the model. As the saying goes, “Data and features determine the upper limit of machine learning, and models and algorithms only approach this upper limit”. Therefore, in the practical training process of traditional machine learning, model performance is largely affected and hindered by data. It is mainly in the following four aspects. First, the complexity of data, which usually contain multi-dimensional, multi-level, and multi-granularity information, makes the application and

processing of data complex and diverse. Second, the heterogeneous data [17], which often contain non-single mixed information, such as numerical and categorical information, make it challenging to process data effectively. Third, the uncertainty [18], redundancy [19], and inconsistency [20] of the data bring certain difficulties to the classification task. Fourth, the information contained in missing data [21] is tough to use effectively.

In order to solve the above problems in telecom fraud, achieve fraud mining, and avoid unnecessary economic losses, a large amount of telecom fraud research has emerged. Traditional telecom fraud detection methods typically rely on compiling blacklists of fraudulent numbers to enable fraudulent user discovery and detection. However, fraudulent strategies have evolved, making traditional methods no longer applicable. Therefore, to mine valuable information for fraud detection from multiple network domains of telecommunication data (SMS data, user data, call communication data, app Internet data), behavioral interaction-based [22], topology-based [23], and content-based [24] approaches arise. Meanwhile, considering the rarity and expensive nature of labeled data, unsupervised methods [25,26] are utilized to achieve fraud mining. However, the above studies lack the consideration of fraud from the perspective of the uncertainty of the data itself. The incompleteness of data or the relevance of attributes plays a critical role in the effective detection of fraud problems. Information theory and rough set theory as valid means of measuring uncertainty provide new ideas for solving the telecommunication fraud problem.

In recent years, with the intensive study of rough set theory [27], outlier detection methods based on rough sets and information theory have received extensive attention and research, which provide theoretical support for discovering important information and classifying complex objects. It has strong interpretability and can deal with unlabeled, heterogeneous, redundant, incomplete, or uncertain data. Attribute reduction [19–21,28–32], or feature selection, is a method to simplify data, reduce data dimension, and improve model classification ability by filtering out irrelevant or redundant features in data, which can effectively avoid overfitting problems. However, vanilla attribute reduction algorithms [33] of classical rough set theory can only learn the information through strict indistinguishable relation division of the data. This equivalence relation is too tough to handle the incomplete, the ordered, the mixed, and the dynamic data, and these algorithms have poor fault tolerance. To overcome this limitation, variants of rough set theory, for example, the attribute importance based [19,20], the positive region based, the tolerance relation based [28], the maximal consistent block based [21], the discernibility matrix based [29], and the incremental based [30] have proved effective in incomplete information systems [34], ordered information systems [35], mixed-valued information systems [14], and dynamic information systems [36]. Generally speaking, the discernibility matrix-based is time-consuming and infeasible for large-scale datasets, while the attribute importance-based has low time complexity. Moreover, tolerance relation is the weakened form of indistinguishable relations, which can validly solve incomplete information. Maximal consistent block describes the maximal objects set under the tolerance relationship, meaning that there is neither redundant, irrelevant information nor information loss. In contrast, the maximal consistent block accurately expresses the objects' information under coverage and has higher accuracy.

After weighing the applicability of these variants, this paper introduces a maximal consistent block to deal with the uncertainty, incompleteness, and redundancy of data in the telecom fraud detection problem for the first time. Guided and inspired by previous research, an anomaly detection method (MCIR-CGAD) based on correlation and the maximal consistent block is proposed in this paper. The main contributions of this paper are summarized as follows:

- From the perspective of improving data quality based on the entropy function under rough set theory, we analyze the effect of attribute correlation and independence on the importance of attributes. A max-correlation and max-independence rate attribute reduction algorithm (MCIR) is designed to eliminate redundancy and noise contained in the data.

- From the perspective of data incompleteness processing, a rough gain anomaly detection algorithm (RGAD) is constructed based on the maximal consistent blocks and information gain, which can effectively supply missing data and provide an effective solution for incomplete data processing and feature information measurement.
- The effectiveness of the MCIR-RGAD algorithm is verified in the UCI dataset and authentic telecom fraud dataset. The results show that compared with the other eight kernel functions, the MCIR-RGAD algorithm can reduce the time complexity and effectively use the information contained in the missing data to improve the model performance.

The remainder of this paper is organized as follows. Section 2 gives the basic preliminaries of rough set theory. The design of the MCIR-RGAD algorithm is proposed in Section 3. Furthermore, Section 4 conducts the experimental analysis, and Section 5 summarizes the conclusions.

## 2. Preliminaries

### 2.1. Rough Set Theory

Rough set theory is an effective way to tackle and utilize incomplete datasets. The information contained in datasets can be represented as an information system.

An information system  $(U, A, V, f)$  is a decision information system, where  $U = \{x_1, x_2, \dots, x_n\}$  is a nonempty finite set of objects known as a universe. Set  $A = C \cup D = \{a_1, a_2, \dots, a_m, D\}$  is composed of the condition attribute set  $C = \{a_1, a_2, \dots, a_m\}$  and the decision attribute set  $D$ , where  $C \cap D = \emptyset$ . The information function  $f : U \times A \rightarrow V$  is a map from the attribute of an object to information value, i.e.,  $f(U, A) = V$ . Normally, a decision information system  $(U, A, V, f)$  can be abbreviated as  $(U, A)$ .

**Definition 1** (Indistinguishable Relation [37]). *Given an information system  $(U, A)$ ,  $A = C \cup D$ ,  $\forall B \subseteq C$  is an attribute subset. An equivalence relation on the set  $U$  is called the indistinguishable relation  $IND(B)$ , if it satisfies:*

$$IND(B) = \{(x, y) \in U \times U \mid \forall a \in B, f(x, a) = f(y, a)\}, \tag{1}$$

where  $[x]_{IND(B)} = \{y \mid (x, y) \in IND(B)\}$  is a set of equivalence relations about  $x$ . Set family  $U/IND(B) = \{[x]_{IND(B)} \mid x \in U\} = \{X_1, X_2, \dots, X_m\}$  means a partition of  $U$  about attribute set  $B$ .  $U = \cup_{i=1}^m X_i$  and  $X_i \cap X_j = \emptyset (i \neq j)$ . Normally,  $[x]_{IND(B)}$  and  $U/IND(B)$  can be abbreviated as  $[x]_B$  and  $U/B$ , respectively.

In an incomplete information system, the indistinguishable relation is unable to effectively divide the incomplete information. Then, the tolerance relation is given as follows.

**Definition 2** (Tolerance Relation [37]). *Given an incomplete information system  $(U, A)$ ,  $A = C \cup D$ .  $\forall B \subseteq C$  is an attribute subset. The binary relation of incomplete information on  $U$  is defined as*

$$SIM(B) = \{(x, y) \in U \times U \mid f(x, a) = f(y, a), \text{ or } f(x, a) = *, \text{ or } f(y, a) = *, \forall a \in B\}, \tag{2}$$

where  $*$  means the incomplete information. Denote  $U/SIM(B)$  as the family of all equivalence classes of  $SIM(B)$ , or simply  $U/B$ .

**Definition 3** (Maximal Consistent Block [31]). *Given an incomplete information system  $(U, A)$ ,  $A = C \cup D$ ,  $B \subseteq C$  is an attribute subset, and  $Y$  is said to be a maximal consistent block of attribute set  $B$ . If  $Y$  satisfies*

- (i)  $\forall x, y \in Y \subseteq U$ , s.t.  $(x, y) \in SIM(B)$ , then  $Y$  is called a consistent block;
- (ii)  $\nexists X \in MCB(B)$ , s.t.  $Y \subseteq X$ .

where  $MCB(\mathcal{B})$  is the set of all maximal consistent blocks with  $\mathcal{B} \subseteq \mathcal{A}, \forall x \in \mathcal{U}$ . The set of all MCB of  $x$  is denoted by  $MCB_x(\mathcal{B})$ , where  $MCB_x(\mathcal{B}) = \{Y \mid Y \in MCB(\mathcal{B}), x \in Y\}$ .

**Example 1.** Consider descriptions of several users of the telecom network in Table 1. It is an incomplete decision information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ , where  $\mathcal{U} = \{x_1, x_2, \dots, x_5\}$ ,  $\mathcal{A} = \{a_1, a_2, a_3\}$  with  $a_1$ -Duration,  $a_2$ -Place,  $a_3$ -Platform, and  $*$  means the incomplete information.

According to the tolerance relation in Definition 2, it follows that  $\mathcal{U}/\mathcal{A} = \{[x_1]_{\mathcal{A}}, \dots, [x_5]_{\mathcal{A}}\}$ , where  $[x_1]_{\mathcal{A}} = [x_5]_{\mathcal{A}} = \{x_1, x_5\}$ ,  $[x_2]_{\mathcal{A}} = \{x_2, x_3\}$ ,  $[x_3]_{\mathcal{A}} = \{x_2, x_3, x_4\}$ ,  $[x_4]_{\mathcal{A}} = \{x_3, x_4\}$ .

By the concept of the maximal consistent block in Definition 3, the maximal consistent block of attribute set  $\mathcal{A}$  is  $MCB(\mathcal{A}) = \{[x_1]_{\mathcal{A}}, [x_2]_{\mathcal{A}}, [x_4]_{\mathcal{A}}\}$ .

**Table 1.** An incomplete information system  $(\mathcal{U}, \mathcal{A})$  about the telecom communication heterogeneous data.

ID	Duration	Amount	Place	Platform	Fraud
$x_1$	$[60, +\infty]$	1	Foreign	WeChat	Yes
$x_2$	$[0, 60]$	*	Foreign	Telecom	No
$x_3$	*	3	*	Telecom	No
$x_4$	*	3	Domestic	Telecom	No
$x_5$	$[60, +\infty]$	1	Foreign	*	Yes

Note:  $*$  means the incomplete information.

**Definition 4** (Information Granularity [37]). Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}, \forall \mathcal{B} \subseteq \mathcal{C}$  is an attribute subset, and the information granularity of attribute  $\mathcal{B}$  is defined as

$$G(\mathcal{B}) = \frac{1}{|\mathcal{U}|} \sum_{i=1}^{|\mathcal{U}|} \frac{|[x_i]_{\mathcal{B}}|}{|\mathcal{U}|}, \tag{3}$$

where  $|[x_i]_{\mathcal{B}}|$  and  $|\mathcal{U}|$  mean the number of the indistinguishable relation set  $[x]_{\mathcal{B}}$  and set  $\mathcal{U}$ , respectively.

**Remark 1.** Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}, \forall \mathcal{B} \subseteq \mathcal{C}$ , conditional granularity, mutual information granularity, and joint granularity of attribute set  $\mathcal{B}$  and  $\mathcal{D}$  are defined as [28,38]  $CG(\mathcal{B} \mid \mathcal{D}) = \sum_{i=1}^{|\mathcal{U}|} \frac{|[x_i]_{\mathcal{B}}| - |[x_i]_{\mathcal{B}} \cap [x_i]_{\mathcal{D}}|}{|\mathcal{U}|^2}$ ,  $MG(\mathcal{B}; \mathcal{D}) = \sum_{i=1}^{|\mathcal{U}|} \frac{|[x_i]_{\mathcal{B}}| + |[x_i]_{\mathcal{D}}| - |[x_i]_{\mathcal{B}} \cap [x_i]_{\mathcal{D}}|}{|\mathcal{U}|^2}$ ,  $JG(\mathcal{B} \cup \mathcal{D}) = \sum_{i=1}^{|\mathcal{U}|} \frac{|[x_i]_{\mathcal{B} \cap \mathcal{D}}|}{|\mathcal{U}|^2}$ , where  $|[x_i]_{\mathcal{B}} \cap [x_i]_{\mathcal{D}}| = |[x_i]_{\mathcal{B} \cup \mathcal{D}}|$  means the division of knowledge under attribute  $\mathcal{B}$  and attribute  $\mathcal{D}$ .

### 2.2. Information Theory

Information entropy is a measure of system uncertainty from the perspective of an information view. The magnitude of entropy reflects the degree of chaos or uncertainty of the system through the distribution of data information.

**Definition 5** (Information Entropy [37]). Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}, \forall \mathcal{B} \subseteq \mathcal{C}$  is an attribute subset, and the information entropy  $H(\mathcal{B})$  is defined as

$$H(\mathcal{B}) = -\frac{1}{|\mathcal{U}|} \sum_{i=1}^{|\mathcal{U}|} \log_2 \left( \frac{|[x_i]_{\mathcal{B}}|}{|\mathcal{U}|} \right), \tag{4}$$

where  $|\mathcal{U}|$  means the element number of object set  $\mathcal{U}$ .

**Remark 2.** By [37], the information entropy is called the granulation measure. The equivalent definition of the complete information system  $(U, A)$  in Equation (4) is defined as

$$H(B) = - \sum_{i=1}^{|U/B|} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|}, \tag{5}$$

where  $U/B = \{[x]_B \mid x \in U\} = \{X_1, X_2, \dots, X_m\}$ ,  $|U/B| = m$ , and  $X_i \cap X_j = \emptyset, i \neq j$ . The form of Equation (5) is consistent with the basic definition of information entropy  $H(X) = - \sum_{i=1}^{|U/B|} p_i \log_2 p_i$ , where  $p_i = \frac{|X_i|}{|U|}, \sum_{i=1}^m p_i = 1$ . Therefore, we can understand the change of information entropy from the relationship between sets by a Venn diagram. In the information theory of rough set theory, the finer the partition, the bigger the entropy.

**Remark 3.** In a complete information system  $(U, A)$ , condition entropy  $H(D \mid B)$ , mutual information  $H(D; B)$ , and joint entropy  $H(D \cup B)$  of attribute  $B$  and  $D$  are defined as [39]

$$H(D \mid B) = - \sum_{i=1}^m \sum_{j=1}^n \frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i \cap Y_j|}{|X_i|}, \quad H(D; B) = - \sum_{i=1}^m \sum_{j=1}^n \frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i| \cdot |Y_j|}{|X_i \cap Y_j| \cdot |U|},$$

$$H(D \cup B) = - \sum_{i=1}^m \sum_{j=1}^n \frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i \cap Y_j|}{|U|}.$$

**Theorem 1 (Entropy Measure).** Given an incomplete information system  $(U, A), A = C \cup D, \forall B \subseteq C$ , conditional entropy, mutual information, and joint entropy of attribute  $B$  and  $D$  are defined as

$$(i) CH(D \mid B) = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[x_i]_B \cap [x_i]_D|}{|[x_i]_B|},$$

$$(ii) MH(D; B) = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[x_i]_B| \cdot |[x_i]_D|}{|[x_i]_B \cap [x_i]_D| \cdot |U|} \tag{6}$$

$$(iii) JH(D \cup B) = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[x_i]_B \cap [x_i]_D|}{|U|},$$

where  $MH(D; B) = H(D) - CH(D \mid B)$ .

**Proof of Theorem 1.** The specific proof of Theorem 1 can be found in Appendix A.  $\square$

### 3. Max-Correlation and Max-Independence Rate-Rough GAIN anomaly Detection Algorithm (MCIR-RGAD)

In the information age of Industry 4.0, the amount of data containing a large number of attributes has proliferated. However, not all attributes are relevant to the classification task. In cyberspace, data may be relevant, repetitive, or similar, which does not bring new and valuable information to the anomaly detection task, leading to unnecessary time costs. In addition, attributes that are not relevant to the anomaly detection task may be noisy and not only fail to help model learning, but may even affect detection performance. In addition, data may inevitably be lost during collection, processing, and storage. The lost data itself may contain hidden anomaly information, and a simple subjective assignment or deletion may lead to an invalid use of the lost information. As a result, the attributes in the data are usually not fully functional.

The entropy function of information theory, as a quantitative paradigm for measuring uncertainty, can effectively measure the correlation between attributes in data. In addition, the missing incomplete data has a certain degree of uncertainty, and this uncertainty may also contain valuable information. Therefore, this paper uses the mutual information function in information theory to measure important attributes that are highly relevant and less redundant to the classification task. Further, the maximal consistent blocks of rough set theory are used to process the missing data, and the information useful for the anomaly detection task is mined from the perspective of uncertainty of incomplete data to realize the improvement of the anomaly detection performance.

The main idea of this section can be divided into three parts. Section 3.1 theoretically discusses the relationship between the attribute correlation and redundancy in the incomplete information system. Then, Section 3.2 presents an optimization algorithm of attribute reduction(MCIR) with a correlation and independent information. Further, we design a rough gain anomaly detection algorithm(RGAD) based on the maximal consistent block to solve the incompleteness of authentic telecom fraud detection in Section 3.3. Figure 1 shows the framework of the proposed methodology.

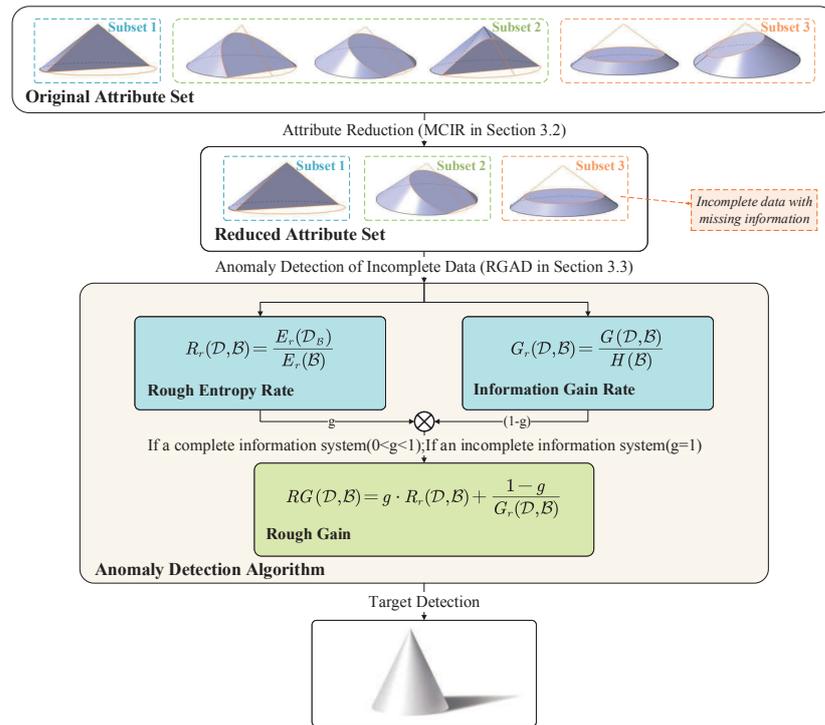


Figure 1. The framework of the proposed methodology.

### 3.1. Relationship of Correlation and Redundancy

To date, many criteria have been proposed to consider the correlation or redundancy of new classification information, such as criteria JMI, CMIM, CIFE, ICAP, ICI, MCI, etc., which are summarized in [19]. The criteria are shown as follows.

$$J_{JMI}(a_j) = \sum_{a_i \in \mathcal{B}} MH(a_i, a_j; \mathcal{D}) \triangleq \sum_i (\textcircled{1}_i + \textcircled{2}_i + \textcircled{3}_i),$$

$$J_{CIFE}(a_j) = MH(\mathcal{D}; a_j) - \sum_{a_i \in \mathcal{B}} MH(\mathcal{D}; a_j; a_i) \triangleq \textcircled{2}_i + \textcircled{3}_i - \sum_i \textcircled{2}_i,$$

$$J_{ICAP}(a_j) = MH(\mathcal{D}; a_j) - \sum_{a_i \in \mathcal{B}} \max MH(\mathcal{D}; a_j; a_i) \triangleq \textcircled{2}_i + \textcircled{3}_i - |\mathcal{B}| \max_i \textcircled{2}_i,$$

$$J_{ICI}(a_j) = MH(\mathcal{D}; a_i | a_j) + MH(\mathcal{D}; a_j | a_i) \triangleq \textcircled{1}_i + \textcircled{3}_i,$$

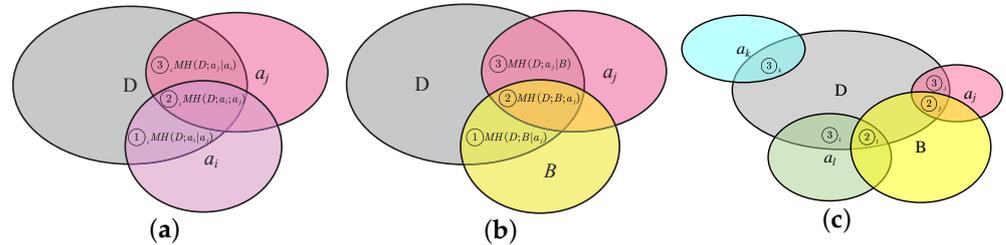
$$J_{MRI}(a_j) = MH(\mathcal{D}, a_j) + \sum_{a_i \in \mathcal{B}} ICI(\mathcal{D}; a_i, a_j) \triangleq \textcircled{2}_i + \textcircled{3}_i + \sum_i (\textcircled{1}_i + \textcircled{3}_i),$$

where  $MH(\mathcal{D}; a_i | a_j) = MH(\mathcal{D}; a_i, a_j) - MH(\mathcal{D}; a_j)$ ,  $\textcircled{1}_i = MH(\mathcal{D}; a_i | a_j)$ ,  $\textcircled{2}_i = MH(\mathcal{D}; a_i; a_j)$ , and  $\textcircled{3}_i = MH(\mathcal{D}; a_j | a_i)$ .

In Figure 2a,  $\textcircled{1}_i$  manifests the relevant information of selected attribute  $a_i \in \mathcal{B}$ ,  $\textcircled{2}_i$  means the redundant information between the attributes  $a_i, a_j$ , and  $\mathcal{D}$ , and  $\textcircled{3}_i$  represents the relevant information of candidate attribute.

In the literature, the correlation and redundancy of the criterion function are frequently compared between each candidate attribute  $a_j$  and each attribute  $a_i$  of the selected attribute set  $\mathcal{B}$ . However, this comparison method has a lot of redundant calculations about informa-

tion. Therefore, this paper regards the selected attribute reduction set  $\mathcal{B}$  as a whole and studies the correlation and redundancy between the candidate attribute  $a_j$  and the attribute reduction set  $\mathcal{B}$ , as shown in Figure 2b.



**Figure 2.** The relationship of relevance and independence in the complete information system: (a)  $a_i$  is a variable,  $a_j$  is fixed; (b)  $a_j$  is a variable,  $B$  is fixed; (c) candidate attribute selection.

For the convenience of formulation, we set ①, ②, and ③ denoted as ① =  $MH(\mathcal{D};\mathcal{B}|a_j)$ , ② =  $MH(\mathcal{D};\mathcal{B};a_j)$ , and ③ =  $MH(\mathcal{D};a_j|\mathcal{B})$ , respectively. In Figure 2b, ① manifests the relevant information of selected attribute  $a_i \in \mathcal{B}$ , ② means the redundant information between the attributes  $\mathcal{B}$ ,  $a_j$ , and  $\mathcal{D}$ , and ③ represents the relevant and independent information of candidate attribute  $a_j$ .

**Theorem 2** (③  $\triangleq$  ① + ② + ③). Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ ,  $\mathcal{B} \subseteq \mathcal{C}$  has been selected, and  $a_j$  is a candidate attribute, then the correlation and redundancy relationship between the attribute  $a_j$ ,  $\mathcal{B}$ , and  $\mathcal{D}$  satisfies

$$MH(\mathcal{D};a_j|\mathcal{B}) \triangleq MH(\mathcal{D};\mathcal{B} \cup a_j). \tag{7}$$

**Proof of Theorem 2.** According to the definition of symbols ①, ②, and ③, we deduce

$$MH(\mathcal{D};\mathcal{B} \cup a_j) - MH(\mathcal{D};a_j|\mathcal{B}) = (\textcircled{1} + \textcircled{2} + \textcircled{3}) - \textcircled{3} = MH(\mathcal{D};\mathcal{B}).$$

In the decision information system, the attribute set  $\mathcal{B}$  has been selected, and  $\mathcal{D}$  is certain, so the division of knowledge is definite. Then,  $MH(\mathcal{D};\mathcal{B})$  is a constant. There is a nonnegative constant  $\Delta_3 = MH(\mathcal{D};\mathcal{B})$ , such that

$$MH(\mathcal{D};\mathcal{B} \cup a_j) - MH(\mathcal{D};a_j|\mathcal{B}) = \Delta_3. \tag{8}$$

Hence, Equation (7) holds, i.e., ③  $\triangleq$  ① + ② + ③.  $\square$

Theorem 2 manifests that the correlation of the newly selected attribute  $a_j$  is consistent with the attribute reduction set  $\mathcal{B} \cup a_j$ , and the effect is the same in classification detection.

**Theorem 3** (① + ③  $\triangleq$  ③ - ②). Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ , suppose attribute set  $\mathcal{B} \subseteq \mathcal{C}$  has been selected, and  $a_j$  is a candidate attribute, then the correlation and redundancy relationship between the attribute  $a_j$ ,  $\mathcal{B}$ , and  $\mathcal{D}$  satisfies

$$MH(\mathcal{D};\mathcal{B}|a_j) + MH(\mathcal{D};a_j|\mathcal{B}) \triangleq MH(\mathcal{D};a_j|\mathcal{B}) - MH(\mathcal{D};\mathcal{B};a_j). \tag{9}$$

**Proof of Theorem 3.** According to the definition of symbols ①, ②, and ③, we have that  $|(MH(\mathcal{D};\mathcal{B}|a_j) + MH(\mathcal{D};a_j|\mathcal{B})) - (MH(\mathcal{D};a_j|\mathcal{B}) - MH(\mathcal{D};\mathcal{B};a_j))| = |(\textcircled{1} + \textcircled{3}) - (\textcircled{3} - \textcircled{2})|$ .

Based on Equation (8) in Theorem 2, ③ +  $\Delta_3$  = ① + ② + ③ holds, hence

$$|(MH(\mathcal{D};\mathcal{B}|a_j) + MH(\mathcal{D};a_j|\mathcal{B})) - (MH(\mathcal{D};a_j|\mathcal{B}) - MH(\mathcal{D};\mathcal{B};a_j))| \tag{10}$$

$$\begin{aligned}
 &= |(\textcircled{1} + \textcircled{3}) - (\textcircled{3} - \textcircled{2})| \\
 &= |(\textcircled{1} + \textcircled{3}) - [(\textcircled{1} + \textcircled{2} + \textcircled{3} - \Delta_3) - \textcircled{2}]| \\
 &= |(\textcircled{1} + \textcircled{3}) - (\textcircled{1} + \textcircled{3} - \Delta_3)| = \Delta_3
 \end{aligned}$$

Hence, Theorem 3 is proved, i.e.,  $\textcircled{1} + \textcircled{3} \triangleq \textcircled{3} - \textcircled{2}$ .  $\square$

Theorem 3 shows that only the correlation between the new attribute  $a_j$  and the selected attribute set  $\mathcal{B}$  is considered, which is equivalent to considering the correlation and redundancy of new attributes  $a_j$ .

### 3.2. Max-Correlation and Max-Independence Rate Algorithm (MCIR)

In light of the above analysis and inspired by the literature [19], the max-correlation and max-independence rate algorithm (MCIR) is introduced as follows.

**Definition 6 (MCIR).** Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ ,  $\mathcal{B} \subseteq \mathcal{C}$ , suppose  $a_i \in \mathcal{B}$  and  $a_j \in \mathcal{C} - \mathcal{B}$ , then the max-correlation and max-independence rate function is presented as

$$a_j = \arg \max_{a_j \in \mathcal{C} - \mathcal{B}} J_{MCIR}(a_j), \tag{11}$$

where  $J_{MCIR}(a_j) = \left\{ \frac{MH(\mathcal{D}; a_j | \mathcal{B})}{H(a_j)} \mid \min_{a_j \in \mathcal{C} - \mathcal{B}} MH(\mathcal{D}; \mathcal{B}; a_j) \right\}$ , i.e.  $J_{MCIR}(a_j) = \{\textcircled{3} / H(a_j) \mid \min \textcircled{2}\}$ .

The principle of the MCIR algorithm is to maximize the correlation and the independence of new classification information and minimize the redundancy between old attributes. The definition of information entropy in rough set theory is from the view of the object attribute information division. The finer the division, the greater the entropy value. Therefore, when the system increases the correlation, it tends to select attributes with more new information.

The attribute reduction algorithm based on max-correlation and max-independence rate is shown in Algorithm 1.

---

#### Algorithm 1 Max-Correlation and Max-Independence Rate (MCIR)

---

**Input:** Information system  $(\mathcal{U}, \mathcal{C} \cup \mathcal{D})$ .

**Output:** An attribute reduction set  $\mathcal{B}$ .

- 1: compute  $MH(\mathcal{C}; \mathcal{D})$
  - 2:  $Feat \leftarrow \mathcal{C}$ ,
  - 3:  $\mathcal{B} \leftarrow \arg \max_{a_j \in Feat} |MH(\mathcal{D}; \mathcal{B}) - MH(\mathcal{D}; \mathcal{B} \cup a_j)|$
  - 4:  $Feat \leftarrow \mathcal{C} - \mathcal{B}$ ,
  - 5: **while**  $J_{MH}(a_j) \geq \theta$  **do**
  - 6:   **for**  $a_j \in Feat$  **do**
  - 7:     **if**  $|Feat| = 0$  **then**
  - 8:        $\mathcal{B} \leftarrow \mathcal{B}$
  - 9:     **else**  $\{|Feat| \neq 0\}$
  - 10:        $a_j \leftarrow \arg \max_{a_j \in Feat} J_{MCIR}(a_j)$
  - 11:        $\mathcal{B} \leftarrow \mathcal{B} + \{a_j\}$
  - 12:        $Feat \leftarrow Feat - \{a_j\}$
  - 13:        $J_{MH}(a_j) \leftarrow |CH(\mathcal{D} | \mathcal{C}) - CH(\mathcal{D} | \mathcal{B})|$
  - 14:     **end if**
  - 15:   **end for**
  - 16: **end while**
-

With the data obtained in the different scenarios, the importance of the correlation and redundancy between attributes exists in diversity. In other words, in the incomplete information system, when the effect of correlation is far bigger than the redundancy, it is more effective to add new information related to the decision attribute. When similar, redundant, and repetitive information causes noise to affect the detection and classification, it is necessary to increase the correlation and reduce the redundancy.

From the relationship of relevance and independence of the MCIR algorithm in Definition 6, Figure 2c satisfies  $\textcircled{3}_j < \textcircled{3}_k = \textcircled{3}_l$ ,  $\textcircled{2}_l = \textcircled{2}_j$ . It shows that the order of attribute importance is  $a_k \succeq a_l \succeq a_j$ , i.e., attribute  $a_k$  is better than attribute  $a_l$ , and attribute  $a_l$  is better than attribute  $a_j$ , which can be sorted correctly by the MCIR algorithm.

### 3.3. Rough Gain Anomaly Detection Algorithm with Max-Correlation and Max-Independence Rate (MCIR-RGAD)

An anomaly detection algorithm (MCIR-RGAD) is designed based on the maximal consistent block horizontally supplementing reduced data. Then, anomaly detection is carried out for the new complemented data. Inspired by the design of information gain in the decision tree, the main idea of the MCIR-RGAD algorithm is to construct a correlation function to measure the ability of attribute classification.

The decision tree, one of the basic classification methods of machine learning, achieves classification tasks by the characteristics of data information. It has fast classification speed, strong interpretability, and readability. Generally, the decision tree learning process consists of feature selection, decision tree generation, and decision tree pruning. In the decision tree, to improve the learning efficiency of the decision tree, the kernel functions, such as information gain, information gain rate, or Gini coefficient, are used to select important features, and then, the decision tree is constructed recursively based on the kernel function. To avoid the occurrence of classification overfitting, we prune the decision tree to balance the model complexity while ensuring the fitting accuracy of the training data.

Both attribute reduction and decision tree work by finding significant features that can classify decision features in information systems. Attribute reduction algorithms can effectively find relevant classification features and achieve effective feature selection. In addition, since there may be intersections in the equivalence class of the object set divided by the maximally consistent block in the incomplete information system, the completeness is not satisfied, i.e.,  $\sum_i p_i \geq 1$ , and there is a negative value when using the information gain for decision learning. Therefore, this paper designs an improved algorithm (MCIR-RGAD) to solve the anomaly detection problem in incomplete systems. Moreover, similar, redundant, repeated, or invalid features are filtered out by reducing. Therefore, this paper does not consider decision pruning.

Frequently, missing data is handled simply by deleting the missing row, filling in zero, filling in one, or filling in the previous data information. However, the explicit deletion or subjective filling of the acquired information will destroy the original data information, so that the missing information cannot be effectively utilized and processed. In an incomplete information system, knowledge can be divided according to the compatibility between available and missing information. This division method not only does not lose the existing data information, but also is more objective. The definition of the kernel function, rough gain  $RG$ , is given below.

**Definition 7** ([40] Rough Entropy). Given an incomplete information system  $(U, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ ,  $\forall \mathcal{B} \subseteq \mathcal{C}$ , rough entropy  $E_r(\mathcal{B})$  is defined as

$$E_r(\mathcal{B}) = - \sum_{i=1}^U \frac{1}{|U|} \log_2 \frac{1}{|[x]_{\mathcal{B}}|}, \quad (12)$$

where rough entropy  $E_r(\mathcal{B})$  satisfies  $E_r(\mathcal{B}) + H(\mathcal{B}) = \log_2 |U|$ .

Inspired by the literature [40], this paper presents a generalized form of the definition of rough entropy for decision making in information division as shown in Definition 8.

**Definition 8** (Decision Rough Entropy). *Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ ,  $\forall \mathcal{B} \subseteq \mathcal{C}$ , the maximal consistent block of attributes  $\mathcal{B}$  and  $\mathcal{D}$  are  $MCB(\mathcal{B}) = \{\mathcal{B}_1, \dots, \mathcal{B}_k\}$ ,  $MCB(\mathcal{D}) = \{\mathcal{D}_1, \dots, \mathcal{D}_m\}$ , then the decision rough entropy  $E_r(\mathcal{D}_B)$  is defined as*

$$E_r(\mathcal{D}_B) = - \sum_{j=1}^m \sum_{i=1}^k \frac{|\mathcal{B}_i|}{|\mathcal{U}|} \log_2 \frac{|\mathcal{B}_i \cap \mathcal{D}_j|}{|\mathcal{B}_i|}. \tag{13}$$

**Definition 9** (Rough Gain). *Given an incomplete information system  $(\mathcal{U}, \mathcal{A})$ ,  $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ ,  $\forall \mathcal{B} \subseteq \mathcal{C}$ , the maximal consistent block of attributes  $\mathcal{B}$  and  $\mathcal{D}$  are  $MCB(\mathcal{B}) = \{\mathcal{B}_1, \dots, \mathcal{B}_k\}$ ,  $MCB(\mathcal{D}) = \{\mathcal{D}_1, \dots, \mathcal{D}_m\}$ , then the rough gain are defined as*

$$RG(\mathcal{D}, \mathcal{B}) = g \cdot R_r(\mathcal{D}, \mathcal{B}) + (1 - g) \cdot \frac{1}{G_r(\mathcal{D}, \mathcal{B})}, \tag{14}$$

where  $g \in [0, 1]$  is a positive constant,  $R_r(\mathcal{D}, \mathcal{B}) = \frac{E_r(\mathcal{D}_B)}{E_r(\mathcal{B})}$  is the rough entropy rate,  $E_r(\mathcal{D}_B)$  is decision rough entropy,  $E_r(\mathcal{B})$  is rough entropy of attribute  $\mathcal{B}$ ,  $G_r(\mathcal{D}, \mathcal{B}) = \frac{G(\mathcal{D}, \mathcal{B})}{H(\mathcal{B})}$  is the information gain rate,  $G(\mathcal{D}, \mathcal{B}) = H(\mathcal{D}) - H(\mathcal{D} | \mathcal{B})$  is the information gain,  $H(\mathcal{D}) = - \sum_{i=1}^m \frac{|\mathcal{D}_i|}{|\mathcal{U}|} \log_2 \frac{|\mathcal{D}_i|}{|\mathcal{U}|}$ , and  $H(\mathcal{D} | \mathcal{B}) = - \sum_{i=1}^k \sum_{j=1}^m \frac{|\mathcal{B}_i \cap \mathcal{D}_j|}{|\mathcal{U}|} \log_2 \frac{|\mathcal{B}_i \cap \mathcal{D}_j|}{|\mathcal{B}_i|}$ .

Therefore, this paper selects features based on the MCIR algorithm, then combines the advantage of the information gain with rough entropy to deal with missing data information. We design an anomaly detection algorithm, MCIR-RGAD algorithm, to achieve the task of anomaly detection. The specific algorithm is shown in Algorithm 2.

Essentially, the MCIR-RGAD algorithm replaces the information gain function of the decision tree with the rough gain function in Definition 9. Contrary to the information gain, a smaller rough gain indicates a better attribute, and the other parts are consistent with the decision tree. Therefore, consistent with the decision tree model, the time complexity of this model is  $O(n^2)$ .

**Algorithm 2** MCIR-RGAD algorithm**Input:** Information system  $(\mathcal{U}, \mathcal{C} \cup \mathcal{D})$ , an attribute reduction set  $\mathcal{B}$ , threshold  $\epsilon > 0$ .**Output:** A decision tree  $T$ 

```

1: compute  $\mathcal{B}$  ▷ Using Algorithm 1
2: compute  $MCB(\mathcal{B}) = \{\mathcal{B}_1, \dots, \mathcal{B}_k\}$ ,  $MCB(\mathcal{D}) = \{\mathcal{D}_1, \dots, \mathcal{D}_m\}$ . ▷ Using Definition 3
3: if  $(\mathcal{U}, \mathcal{C} \cup \mathcal{D})$  is incomplete then
4:    $g \leftarrow 1$ ,  $RG(\mathcal{D}, \mathcal{B}) \leftarrow R_r(\mathcal{D}, \mathcal{B})$  (Equation (14))
5: else
6:    $0 < g < 1$ ,  $RG(\mathcal{D}, \mathcal{B}) \leftarrow g \cdot R_r(\mathcal{D}, \mathcal{B}) + (1 - g) \cdot \frac{1}{G_r(\mathcal{D}, \mathcal{B})}$  (Equation (14))
7: end if ▷ Recursion Point
8: for  $a_j \in \mathcal{B}$  do
9:    $a_j \leftarrow \arg \min_{a_j \in \mathcal{B}} RG(\mathcal{D}, a_j)$ 
10:   $\mathcal{B} \leftarrow \mathcal{B} - a_j$ 
11:  if  $RG(\mathcal{D}, a_j) > \epsilon$  then
12:     $Label_T \leftarrow \arg \max_{f(\mathcal{U}, \mathcal{D}_i)} |\mathcal{D}_i|$ 
13:  else
14:     $MCB(a_j) \leftarrow \{\mathcal{X}_1, \dots, \mathcal{X}_s\}$ 
15:    for  $\mathcal{X}_i \in MCB(a_j)$  do
16:       $(\mathcal{U}, \mathcal{C} \cup \mathcal{D}) \leftarrow (\mathcal{X}_i, \mathcal{B} \cup \mathcal{D})$ 
17:      if  $|MCB(\mathcal{D})| = 1$  then
18:         $Label_T \leftarrow f(\mathcal{U}, \mathcal{D})$ 
19:      else if  $\mathcal{B} = \phi$  then
20:         $Label_T \leftarrow \arg \max_{f(\mathcal{U}, \mathcal{D}_i)} |\mathcal{D}_i|$ 
21:      else
22:        return Step 8
23:      end if
24:    end for
25:  end if
26: end for
27: return  $T$ 

```

**4. Experimental Analysis**

The UCI Machine Learning Repository datasets (<https://archive.ics.uci.edu/ml/index.php> accessed on 12 April 2022) and the Sichuan telecom fraud phone datasets (<https://aistudio.baidu.com/aistudio/datasetdetail/40690> accessed on 12 April 2022) are used to verify the effectiveness of the proposed method in this section. The MCIR-RGAD algorithms are coded in Python using Visual Studio Code and were run on a remote server with a GPU, NVIDIA GeForce RTX 3090, 48 RAM.

The Sichuan telecom fraud phone dataset consists of four datasets, namely call data (VOC), short message service data (SMS), user information data (USER), and Internet behavior data (APP). The Union data is an integrated dataset combined based on user phone numbers and contains the attribute of four datasets: Voc dataset, APP dataset, SMS dataset, and User dataset. The details of the datasets are described in Table 2.

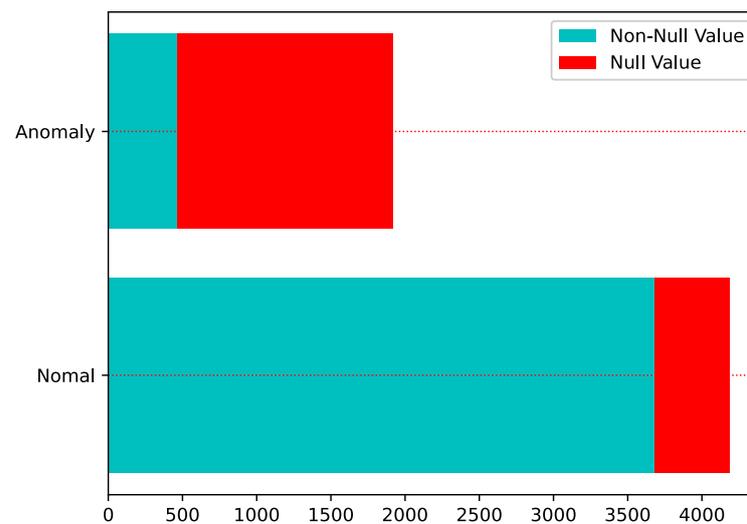
The goal of this paper is to detect fraudulent users among regular users according to the important attribute efficiently selected by the correlation and independence from the perspective of data uncertainty and incompleteness. Next, we discuss the effectiveness of the method proposed in this paper from three aspects: incompleteness of data (Definition 3), MCIR attribute reduction algorithm (Section 3.2), and MCIR-RGAD anomaly detection classifier (Section 3.3).

**Table 2.** Description of datasets.

Datasets	Sample	Attribute	Source
Car	1728	6 + 1	UCI
Adult	48,842	14 + 1	UCI
Bank	4521	16 + 1	UCI
Mushroom	8124	22 + 1	UCI
USER	6106	11 + 1	Telecom
SMS	6,848,509	11 + 1	Telecom
APP	3,283,602	20 + 1	Telecom
VOC	5,015,430	42 + 1	Telecom
Union	6106	84 + 1	Telecom

#### 4.1. Incompleteness of Data

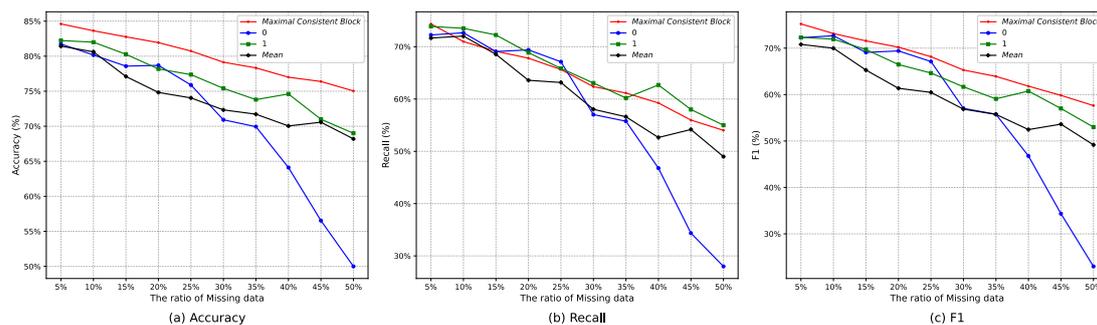
Loss of data during recording, storage, or transmission is a very likely problem. Normally, the way to deal with incomplete information is to delete it directly, fill it with zero, one or mean value; however, this simple way of dealing with it will cause the loss of information. As can be seen from Figure 3, in the Sichuan Telecom fraud dataset, most of the users with null values (red parts) are abnormal, and if they are directly deleted or simply assigned, the abnormal information will not be effectively used. Therefore, from the perspective of improving data quality, this paper uses the idea of maximal consistent blocks in rough set theory to deal with incomplete data to achieve effective information mining.

**Figure 3.** Incomplete data among telecom fraud users.

Then, Table 3 and Figure 4 further illustrate the effectiveness of the maximal consistent block in handling incomplete data. Table 3 and Figure 4 are the performance comparisons of tackling null values under authentic incomplete telecom fraud data and random deletion of artificially constructed incomplete data (5%, 10%, . . . , 50%, 10 types of data missing ratios). From the perspective of the accuracy (Figure 4a), recall (Figure 4b), F1 (Figure 4c), and the number of correct predictions, the maximal consistent block (MCB) can effectively utilize incomplete information and avoid unnecessary information loss.

**Table 3.** Incomplete information processing of authentic telecom fraud data.

Method	Accuracy	Recall	F1	Current Object	Correct Prediction
Drop	86.82%	10.38%	16.42%	4248	3688
Fill 0	78.72%	53.65%	62.10%	6106	4806
Fill 1	78.72%	53.65%	62.10%	6106	4806
Fill Mean	75.86%	48.61%	56.68%	6106	4630
MCB	84.12%	62.47%	71.88%	6106	5136



**Figure 4.** Performance comparison of randomly deleting missing data of the Union dataset under the RGAD algorithm.

4.2. Attribute Reduction under MCIR Algorithm

This paper proposes an attribute reduction algorithm of MCIR which uses the entropy function to measure the correlation and independence of attributes from the perspective of rough set theory. The calculation time of the algorithm is reduced while ensuring the accuracy of the telecom fraud detection problem. The main idea is to reduce the computation time by filtering out partial attributes that are most relevant to fraudulent users and have the greatest independence (least redundancy).

Experiments on UCI and telecom fraud data show that the computation time of the data can be significantly reduced by filtering out important attributes. Figures 5 and 6 further illustrate that the MCIR algorithm not only effectively reduces the computation time, but also eliminates the adverse effects of noise on information, improves data quality, and maintains or even improves the accuracy of model detection.

Generally, datasets can be roughly divided into four types, namely: non-redundant and noise-free dataset (Figure 7a Car, approximated as a strictly monotonically increasing function), non-redundant and noisy dataset (Figure 7b Adult, approximately concave function), redundant and noisy dataset (Figure 7c Bank, approximately non-increasing function), and redundant and non-noise dataset (Figure 7d Mushroom, approximately non-decreasing function). Redundancy shows the approximation, repetition, and correlation of attributes in the data with each other; noise refers to the interference and misleading effects of certain attributes in the data on the classification task. Specifically, for a non-redundant and noise-free dataset, there is no need to perform attribute reduction, and each dimension of features is important information. For other types of data, it is necessary to remove redundant and noisy attributes. In addition, it can be seen from Figure 7 that compared with other different attribute reduction algorithms ( $J_{CG}$  [41],  $J_{CH}$  [42],  $J_{MG}$ ,  $J_{MH}$  [43],  $J_{JG}$ ,  $J_{JH}$ , ③, ① + ② + ③ [44], ① + ③ [19], ③–② [45]), the MCIR algorithm (red dotted line) designed in this paper achieves better accuracy with fewer attributes. Since the MCIR model removes as many redundant or noisy attributes from the data as possible and achieves data optimization through data dimensionality reduction, making the reduced data better for anomaly detection tasks, the model can maintain or even improve the accuracy of performance detection while reducing the time complexity.

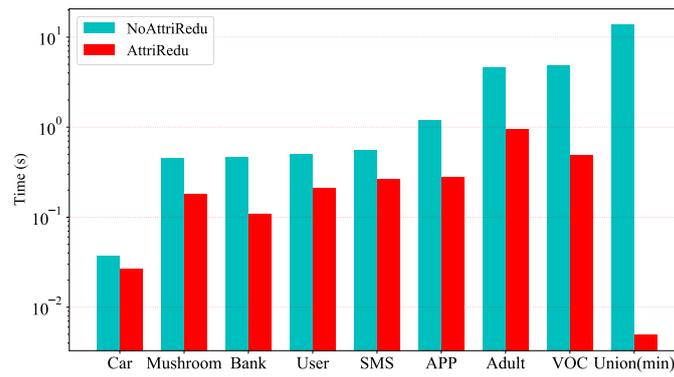


Figure 5. Comparison of the computation time before and after the MCIR-RGAD algorithm.

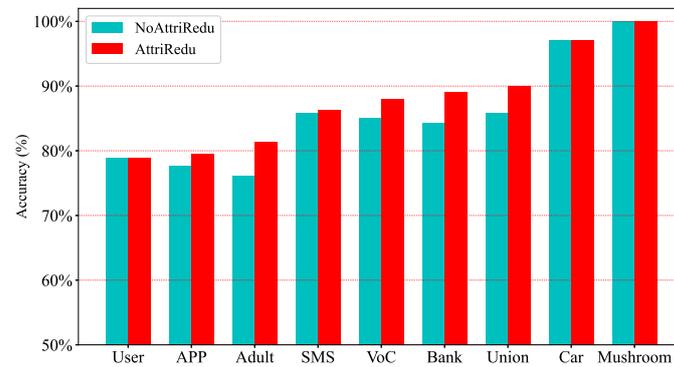


Figure 6. Comparison of the classification accuracy before and after the MCIR-RGAD algorithm.

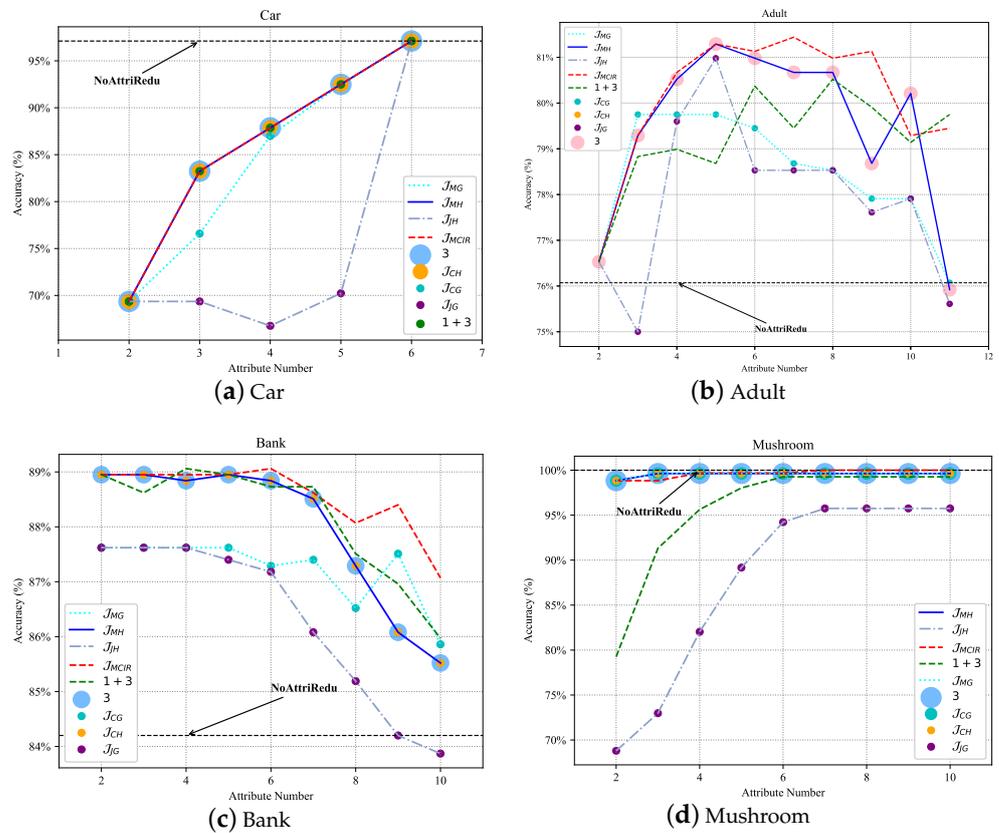


Figure 7. Performance comparison of different attribute reduction algorithms under the MCIR-RGAD algorithm.

Therefore, in the process of data processing, the MCIR algorithm can use partial important attribute information to shorten the computation time and effectively improve the detection accuracy of the model (Figure 7, the black dotted line).

Next, the feature selection of the telecom fraud dataset under the MCIR algorithm is discussed. Figure 8 shows correlations within attributes via a heatmap. Among them, Figure 8a,b are the correlations before and after attribute reduction, respectively. As shown in Figure 8, when attribute reduction is not performed, the data contain a lot of redundant information (dark patches). This paper constructs the MCIR attribute reduction algorithm from the perspective of attribute uncertainty and correlation, which can reduce the information redundancy degree of data while reducing data weight.

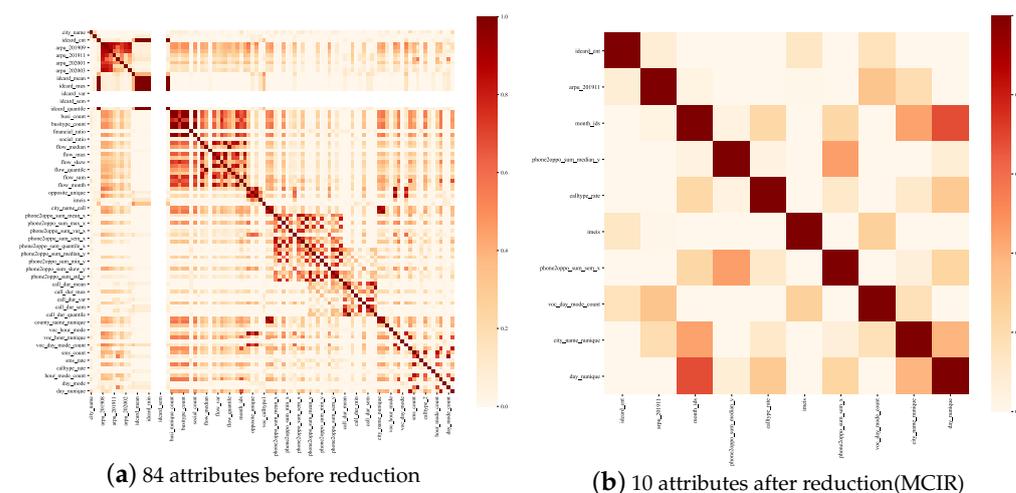
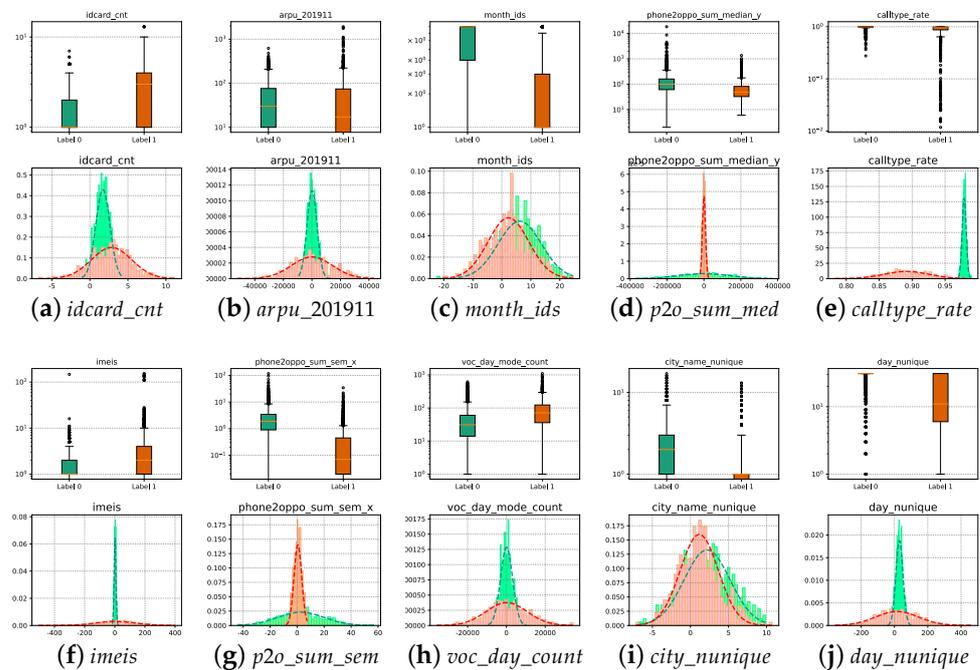


Figure 8. Attribute correlations in the Union dataset.

Further, the boxplot and probability distribution plot in Figure 9 show the difference in statistical distribution between normal and abnormal users. The important attributes selected based on the MCIR-RGAD algorithm can effectively highlight the difference between abnormal users and normal users, and fraudulent users can be filtered out by the selected important attributes. Compared with the original Union dataset of 84 attributes with 85.84% detection performance (Table 4), the detection performance of 10 attributes after MCIR simplification is improved to 89.96%, indicating that the MCIR model involved in this paper effectively achieves the selection of important attributes. To further visualize how the selected attributes distinguish between normal and fraudulent users, Figure 9 depicts the box line plot and statistical distribution of the 10 important attributes in the telecom fraud dataset filtered by the MCIR method. From Figure 9, it can be seen that the distributions of normal users and fraudulent users under the 10 attributes have large differences, mainly in the form of (a, f, e) with large difference in mean and variance, (b, d, j, g, h) with large difference in variances with similar means, and (c, i) with large difference in means with similar variances. The larger the difference between the mean and variance distributions of normal and fraudulent users for the selected attributes, the more effective it the method is in distinguishing fraudulent users.

**Table 4.** Comparison of accuracy, computational time, and robustness of attribute reduction in the MCIR-RGAD algorithm.

Datasets	Attribute Number		Accuracy(%)		Time(s)		Robustness(%)
	Before	After	Before	After	Before	After	MCIR-RGAD
Car	6	5	97.11	92.49	0.0373	0.0254	69.91
Adult	14	5	76.07	81.29	0.2765	0.0517	96.64
Bank	16	5	84.20	89.06	0.4618	0.1074	94.16
Mushroom	22	7	100.00	100.00	0.4523	0.1816	83.94
User	11	7	78.81	78.81	0.4090	0.2223	78.26
SMS	11	7	85.75	86.24	0.5548	0.2196	84.51
APP	20	7	77.66	79.54	0.9432	0.3015	88.67
Voc	42	12	84.98	88.05	4.8257	0.4677	97.99
Union	84	10	85.81	89.96	830.2283	0.3520	102.88



**Figure 9.** Classification of selected attributes under the MCIR-RGAD algorithm.

### 4.3. Anomaly Detection under MCIR-RGAD Algorithm

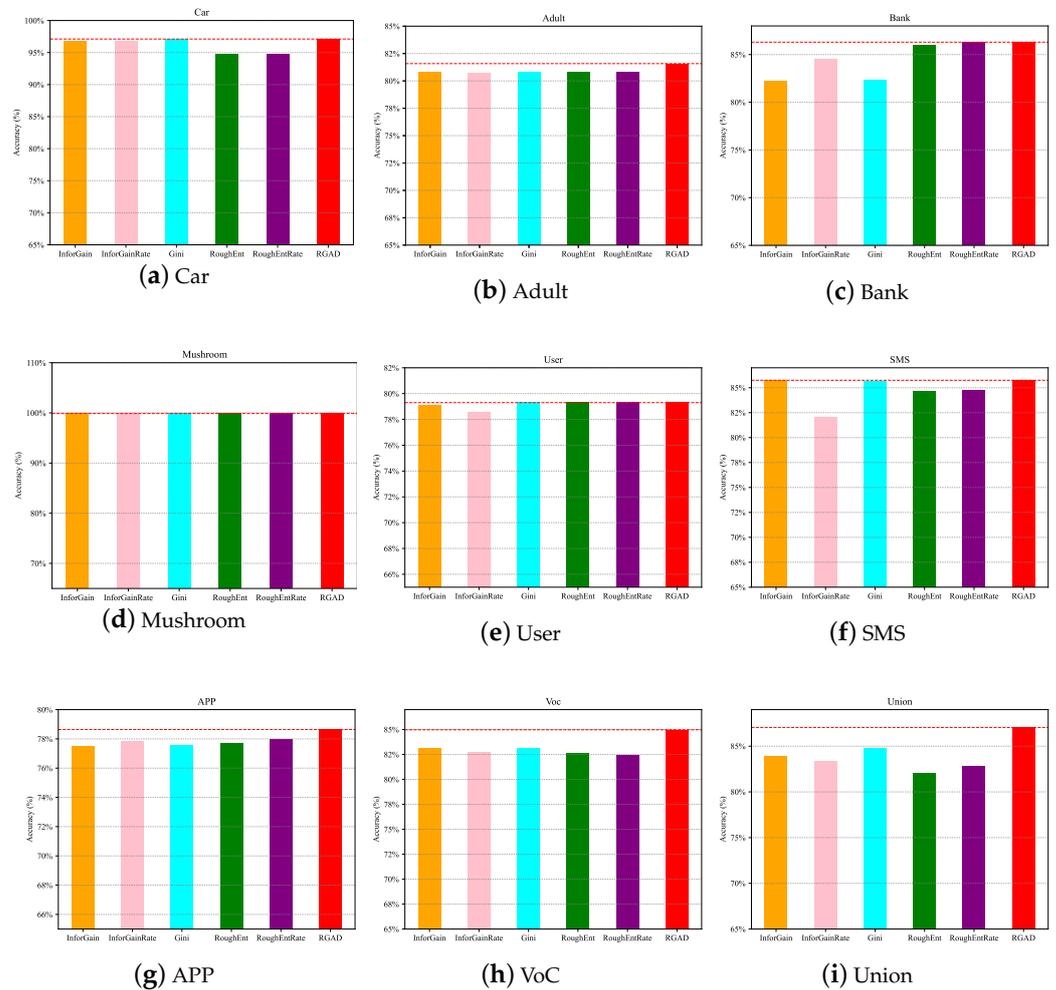
Redundancy and noise attributes are removed from the original data to improve the data quality of the MCIR algorithm. Then, to perform effective anomaly detection on incomplete data containing missing content, this paper designs the MCIR-RGAD algorithm based on maximal consistent blocks. It provides an effective solution for the processing and utilization of incomplete data.

In the anomaly detection of the decision tree, six types of kernel function classification algorithms, namely Information Gain  $G(\mathcal{D}, \mathcal{B})$ , Information Gain Rate  $G_r(\mathcal{D}, \mathcal{B})$ , Gini Coefficient, Rough Entropy  $R_r(\mathcal{D}, \mathcal{B})$ , Rough Entropy Rate  $R_r(\mathcal{D}, \mathcal{B})$ , and Rough Gain  $RG(\mathcal{D}, \mathcal{B})$ , are compared in this paper. As shown in Figure 10, the rough gain anomaly detection algorithm (RGAD) integrates rough entropy and information gain as the kernel function has better performance.

The performance and computation time of nine types of attribute reduction algorithms are shown in Tables 5 and 6. Compared with other algorithms, the MCIR-RGAD algorithm proposed in this paper can effectively achieve classification detection in a shorter time.

**Table 5.** Comparison of classification accuracies of attribute reduction from set view and information view.

Datasets	Set Theory View			Information Theory View			Independence		
	$J_{CG}$	$J_{MG}$	$J_{JG}$	$J_{CH}$	$J_{MH}$	$J_{JH}$	③	① + ③	$J_{MCIR-RGAD}$
Car	92.49%	92.49%	70.23%	92.49%	92.49%	70.23%	92.49%	92.49%	<b>92.49%</b>
Adult	79.75%	79.75%	80.98%	81.29%	81.29%	80.98%	81.29%	78.68%	<b>81.29%</b>
Bank	87.29%	87.29%	87.18%	88.84%	88.84%	87.18%	88.84%	88.73%	<b>89.06%</b>
Mushroom	99.63%	99.63%	95.75%	99.63%	99.63%	95.75%	99.63%	99.26%	<b>100.00%</b>
User	78.81%	78.81%	78.81%	78.81%	78.81%	78.40%	78.81%	78.81%	<b>78.81%</b>
SMS	85.50%	85.50%	80.92%	86.24%	86.24%	80.92%	86.24%	85.59%	<b>86.24%</b>
APP	79.87%	79.87%	79.87%	79.54%	79.54%	<b>79.87%</b>	79.54%	79.38%	79.54%
VoC	85.15%	85.15%	83.32%	87.97%	87.97%	83.32%	87.97%	82.41%	<b>88.05%</b>
Union	88.22%	88.22%	75.44%	88.71%	88.71%	75.68%	88.71%	85.98%	<b>89.96%</b>



**Figure 10.** Performance comparison of six classification detection algorithms.

**Table 6.** Comparison of classification computation time of attribute reduction from set view and information view.

Datasets	Set Theory View			Information Theory View			Independence		
	$J_{CG}$	$J_{MG}$	$J_{JG}$	$J_{CH}$	$J_{MH}$	$J_{JH}$	③	① + ③	$J_{MCIR-RGAD}$
Car	0.0280	0.0280	0.0329	0.0270	0.0270	0.0320	0.0270	0.0264	<b>0.0254</b>
Adult	0.0689	0.0689	0.0689	0.0569	0.0569	0.0616	0.0569	<b>0.0471</b>	0.0517
Bank	0.1737	0.1737	0.1476	0.1715	0.1715	0.1776	0.1715	0.1785	<b>0.1074</b>
Mushroom	0.1970	0.1970	0.2817	0.1930	0.1930	0.2213	0.1930	0.1941	<b>0.1816</b>
User	0.2559	0.2559	<b>0.2051</b>	0.2300	0.2300	0.2112	0.2300	0.2472	0.2223
SMS	0.2042	0.2042	0.2113	0.2405	0.2405	<b>0.2002</b>	0.2405	0.2378	0.2196
APP	0.2909	0.2909	0.2943	0.2531	0.2531	0.2712	<b>0.2531</b>	0.3134	0.3015
VoC	0.4692	0.4692	<b>0.4321</b>	0.5054	0.5054	0.4997	0.5054	0.4570	0.4677
Union	0.2937	0.2937	0.2975	0.3414	0.3414	0.3016	0.3414	<b>0.2844</b>	0.3520

To effectively measure the trade-off between detection performance and computation time cost of an algorithm, this paper designs a robustness metric in Definition 10. In the robustness metric, since computation time and performance level have different importance in different application scenarios, a linear parameter  $k$  is designed to trade off the importance of time and performance. The telecom fraud problem in this paper pays more attention to the accuracy of the model; hence, the hyperparameter weight in the robustness metric is set as  $k = 0.4$ .

**Definition 10** (Performance Robustness).

$$Robust = k \cdot T + (1 - k) \cdot P, \quad (15)$$

where  $P = \frac{P_a}{P_b}$  is the degree of performance retention,  $T = \frac{T_b - T_a}{T_b}$  is the degree of time optimization,  $P_b$ ,  $P_a$ ,  $T_b$ , and  $T_a$  are the performance and time before and after attribute reduction, respectively, and  $k \in [0, 1]$  is a weight parameter of time, which means the importance of time cost.

Then, Table 4 shows the number of attributes after attribute reduction for different datasets and shows the changes in performance and computation time of the MCIR-RGAD algorithm before and after attribute reduction. Note that this paper compares the performance and computation time of different algorithms in the same number of attribute reduction sets  $\mathcal{B}$ .

The performance robustness metric with less computation time and high performance indicates that the designed classifier algorithm is better. The accuracies and computation time in Tables 5–4 and Figure 11 show the robustness under the different attribute reduction algorithms. Compared with other algorithms, the MCIR-RGAD algorithm has strong robustness. That is, when the number of attributes in the attribute set is reduced to the same number, the anomaly detection algorithm MCIR-RGAD can effectively ensure the accuracy of classification detection while shortening the calculation time.

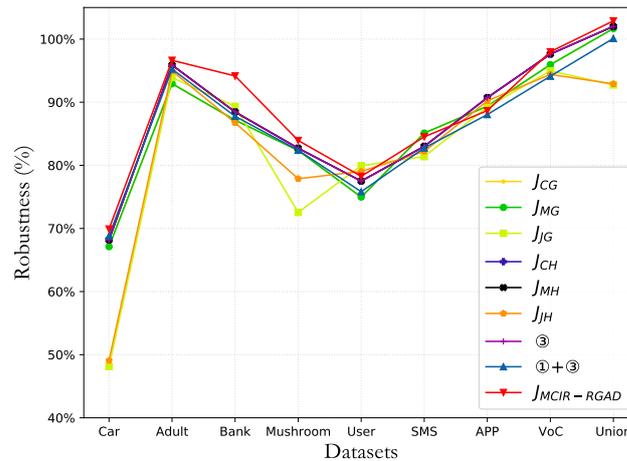


Figure 11. Comparison of the robustness in different datasets.

#### 4.4. Statistical Test Analysis

Two nonparametric statistical test analyses of the Friedman test and the Nemenyi post hoc test are introduced to further verify the validity of the comparison method and the proposed method. We compare the performance differences at a significance level of  $\alpha = 0.05$ .

##### 4.4.1. Friedman Test

The Friedman test can effectively determine whether there is a significant difference in algorithm performance. Suppose we compare  $K$  algorithms on  $N$  datasets. In the Friedman test, the null hypothesis assumes that there is no significant difference between the models. First, the models were ranked on different datasets using the performance accuracy cases in Table 5. Then, we acquire the average of overall ranking for each model,  $R_{ave}^j = \frac{1}{N} \sum_{i=1}^K r_i$ . The performance ranking of the nine algorithms on the nine datasets is given in Table 7. When the performance of the algorithms is equal, the ordinal values are averaged. For example, if the performance of the 7 algorithms ( $J_{CG}, J_{MG}, J_{CH}, J_{MH}, \textcircled{3}, \textcircled{1}+\textcircled{3}, J_{MCIR-RGAD}$ ) under the Car dataset in Table 5 is equal, then their rank values are  $r_i = \frac{1+2+3+4+5+6+7}{7} = 4$ .

The Friedman statistic  $\tau_{\chi^2} = \frac{12N}{K(K+1)} \left( \sum_{i=1}^K r_i^2 - \frac{K(K+1)^2}{4} \right)$  is distributed according to  $\chi^2$ -distribution with  $K - 1$  degrees of freedom, when  $K$  and  $N$  are large enough. Owing to the overly conservative nature of the original Friedman test, the variable  $\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(K-1)-\tau_{\chi^2}}$  is commonly used today, which is distributed according to F-distribution with  $K - 1$  and  $(K - 1)(N - 1)$  degrees of freedom, i.e.,  $\tau_F \sim F(K - 1, (K - 1)(N - 1))$ .

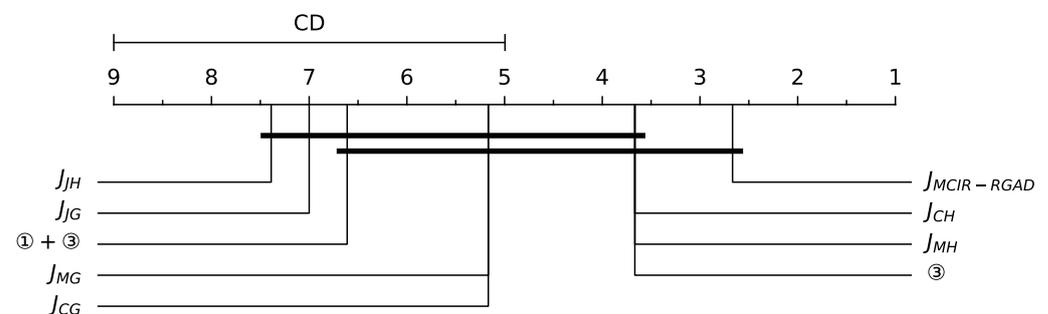
This paper compares nine algorithms using nine datasets. In the Friedman test, if the p-value is less than the significance level or the  $\tau_F$  value is greater than the critical value  $F(8, 64)$  determined by the F-distribution table, the null hypothesis can be rejected, and at least two algorithms are considered to have significant differences. By checking the table and calculating, we have  $\tau_F = 5.0208 > F(8, 64) = 2.0868$  and  $p = 2.688 \times 10^{-5} < 0.05$ . Therefore, the null hypothesis can be rejected with 95% confidence level, indicating that there is a significant difference between the algorithms in the model. Then, a pairwise comparison of the benchmark algorithms was performed using the Nemenyi post hoc test.

**Table 7.** Ranking on 9 datasets for 9 algorithms.

Datasets	Ranking Value $r_i$ (Accuracy(%))								
	$J_{CG}$	$J_{MG}$	$J_{JG}$	$J_{CH}$	$J_{MH}$	$J_{JH}$	① + ③	③	$J_{MCIR-RGAD}$
Car	4 (92.49)	4 (92.49)	8.5 (70.23)	4 (92.49)	4(92.49)	8.5 (70.23)	4 (92.49)	4 (92.49)	4 (92.49)
Adult	7.5 (79.75)	7.5 (79.75)	5.5 (80.98)	2.5 (81.29)	2.5 (81.29)	5.5 (80.98 )	2.5 (81.29)	9 (78.68)	2.5 (81.29)
Bank	6.5 (87.29)	6.5 (87.29)	8.5 (87.18)	3 (88.84)	3 (88.84)	8.5 (87.18)	3 (88.84)	5 (88.73)	1 (89.06)
Mushroom	4 (99.63)	4 (99.63)	8.5 (95.75)	4 (99.63)	4 (99.63)	8.5 (95.75)	4 (99.63)	7 (99.26)	1 (100.00)
User	4.5 (78.81)	4.5 (78.81)	4.5 (78.81)	4.5 (78.81)	4.5 (78.81)	9 (78.40)	4.5 (78.81)	4.5 (78.81)	4.5 (78.81)
SMS	6.5 (85.50)	6.5 (85.50)	8.5 (80.92)	2.5 (86.24)	2.5 (86.24)	8.5 (80.92)	2.5 (86.24)	5 (85.59)	2.5 (86.24)
APP	2.5 (79.87)	2.5 (79.87)	2.5 (79.87)	6.5 (79.54)	6.5 (79.54)	2.5 (79.87)	6.5 (79.54)	9 (79.38)	6.5 (79.54)
VoC	5.5 (85.15)	5.5 (85.15)	7.5 (83.32)	3 (87.97)	3 (87.97)	7.5 (83.32)	3 (87.97)	9 (82.41)	1 (88.05)
Union	5.5 (88.22)	5.5 (88.22)	9 (75.44)	3 (88.71)	3 (88.71)	8 (75.68)	3 (88.71)	7 (85.98)	1 (89.96)
$R_{ave}^j$	5.167	5.167	7	3.667	3.667	7.389	3.667	6.611	2.667

4.4.2. Nemenyi Post Hoc Test

In the Nemenyi post hoc test, the performance of two models is considered to be significantly different if the average rank value  $R_{ave}^j$  of the two models is greater than or equal to the criterion distance ( $CD = q_\alpha \sqrt{\frac{K(K+1)}{6N}}$ ), where the critical value  $q_\alpha$  obeys the Tukey distribution. By checking the table and calculating,  $q_{0.05} = 3.102$  under the confidence level  $\alpha = 0.05$ , then  $CD = 4.0047$ . It can be seen from Figure 12 that the MCIR-RGAD model is optimal and significantly different from  $J_{JH}$  and  $J_{JG}$ . In addition,  $J_{CH}$ ,  $J_{MH}$ , and ③ are equivalent, and  $J_{MG}$  is equivalent to  $J_{CG}$ . Namely, the model performance can be ordered as  $J_{MCIR-RGAD} > J_{CH} = J_{MH} = \textcircled{3} > J_{MG} = J_{CG} > \textcircled{1} + \textcircled{3} > J_{JG} > J_{JH}$ .



**Figure 12.** Average ranks diagram comparing the benchmark methods in terms of accuracy.

5. Conclusions

It is crucial and time-consuming to obtain anomaly classification information in big data with uncertainty, redundancy, and incompleteness. In this paper, a new attribute reduction algorithm (MCIR) is proposed based on the correlation and independence of the data. Furthermore, considering the consistency of attribute reduction and decision tree in selecting features, this paper combines their advantages and constructs an anomaly detection algorithm called RGAD to tackle incomplete data based on the maximal consistent blocks. The proposed algorithm (MCIR-RGAD) can significantly reduce the computation time and effectively maintain or improve the accuracy. Therefore, facing the problem of anomaly detection, this paper provides an effective solution for the optimization of data quality and the processing of incomplete data.

In the future, we plan to extend this work in the context of unsupervised learning from the perspective of structural information among objects, using the concept of neighborhood information systems in rough set theory. The extended work will optimize the data quality and reduce the time complexity through attribute reduction methods, improve the detection performance of classification tasks through structural information, and maximize valuable

information through incomplete mixed data (both categorical and numerical data). This will provide an effective solution to the research of information theory and rough set theory on anomaly detection problems.

**Author Contributions:** Conceptualization, software, R.L. and B.W.; methodology, H.C.; validation, R.L., B.W. and X.H.; formal analysis, K.W.; writing—original draft preparation, R.L.; writing—review and editing, S.L. and H.C.; visualization, X.H.; supervision, project administration, funding acquisition, H.C. and S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Scientific and Technological Special Project of Henan Province under Grant 221100210700.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets used in research are publicly available at <https://archive.ics.uci.edu/ml/index.php> accessed on 12 April 2022 and <https://aistudio.baidu.com/aistudio/datasetdetail/40690> accessed on 12 April 2022. The codes of this research will be made available on request.

**Acknowledgments:** Thanks to the Sichuan Provincial Data Center and Sichuan Mobile for providing the data without which this work would not have been possible. In addition, the authors thank the editors and reviewers for their valuable comments. This work was supported by the by the Major Scientific and Technological Special Project of Henan Province under Grant 221100210700. The authors gratefully thank the associate editor and the reviewers.

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

### Appendix A

**Proof of Theorem 1.** For a complete information system, we have that  $X_i \cap X_j = \emptyset$ ,  $Y_i \cap Y_j = \emptyset, \forall i \neq j$ . Support that  $U/A = \{[x_1]_A, \dots, [x_{|U|}]_A\}$ ,  $X_i = \{x_i^1, x_i^2, \dots, x_i^{k_i}\}$ ,  $Y_j = \{x_j^1, x_j^2, \dots, x_j^{l_j}\}$ , and  $X_i \cap Y_j = \{x_{ij}^1, x_{ij}^2, \dots, x_{ij}^{z_{ij}}\}$ ,  $i = 1, 2, \dots, m, j = 1, 2, \dots, n$  in set  $X_i$ , where  $k_i, l_j$  and  $z_{ij}$  represent the number of elements in the set  $X_i, Y_j$  and  $X_i \cap Y_j$ , respectively.  $X_i \cap Y_j \subseteq X_i$  and  $\bigcup_{j=1}^n (X_i \cap Y_j) = X_i, \bigcup_{i=1}^m (X_i \cap Y_j) = Y_j$ , then  $\sum_{j=1}^n z_{ij} = k_i, \sum_{i=1}^m z_{ij} = l_j, \sum_{i=1}^m X_i = \sum_{i=1}^m k_i = \sum_{j=1}^n Y_j = \sum_{j=1}^n l_j = |U|$ .

$$(i) \text{ Proof of } CH(\mathcal{D} | \mathcal{B}) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[x_i]_{\mathcal{B}} \cap [x_i]_{\mathcal{D}}|}{|[x_i]_{\mathcal{B}}|}.$$

In incomplete information system,  $X_i = [x_i^1]_{\mathcal{B}} = [x_i^2]_{\mathcal{B}} = \dots = [x_i^{k_i}]_{\mathcal{B}}, X_i \cap Y_j = [x_{ij}^1]_{\mathcal{B} \cup \mathcal{D}} = [x_{ij}^2]_{\mathcal{B} \cup \mathcal{D}} = \dots = [x_{ij}^{z_{ij}}]_{\mathcal{B} \cup \mathcal{D}}$ , and  $|X_i| = |[x_i^1]_{\mathcal{B}}| = |[x_i^2]_{\mathcal{B}}| = \dots = |[x_i^{k_i}]_{\mathcal{B}}|, |X_i \cap Y_j| = |[x_{ij}^1]_{\mathcal{B} \cup \mathcal{D}}| = |[x_{ij}^2]_{\mathcal{B} \cup \mathcal{D}}| = \dots = |[x_{ij}^{z_{ij}}]_{\mathcal{B} \cup \mathcal{D}}|$ .

$$\frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i \cap Y_j|}{|X_i|} = \frac{1}{|U|} \left( \log_2 \frac{|[x_{ij}^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} + \dots + \log_2 \frac{|[x_{ij}^{z_{ij}}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} \right) \quad (A1)$$

Then, we have that

$$\begin{aligned}
 & \sum_{j=1}^n \frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i \cap Y_j|}{|X_i|} \\
 &= \sum_{j=1}^n \left( \frac{1}{|U|} \log_2 \frac{|[x_{ij}^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} \right) + \dots + \sum_{j=1}^n \left( \frac{1}{|U|} \log_2 \frac{|[x_{ij}^{z_{ij}}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} \right) \\
 &= \left( \frac{1}{|U|} \log_2 \frac{|[x_{i1}^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} + \dots + \frac{1}{|U|} \log_2 \frac{|[x_{i1}^{z_{i1}}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} \right) \\
 &+ \dots \\
 &+ \left( \frac{1}{|U|} \log_2 \frac{|[x_{im}^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} + \dots + \frac{1}{|U|} \log_2 \frac{|[x_{im}^{z_{im}}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} \right)
 \end{aligned} \tag{A2}$$

According to  $z_{i1} + z_{i2} + \dots + z_{im} = k_i$  and  $\sum_{i=1}^m \sum_{j=1}^n z_{ij} = \sum_{i=1}^m k_i = |U|$ , then  $X_i \cap Y_1 = \{x_{i1}^1, x_{i1}^2, \dots, x_{i1}^{z_{i1}}\}$ ,  $X_i \cap Y_2 = \{x_{i2}^1, x_{i2}^2, \dots, x_{i2}^{z_{i2}}\}, \dots, X_i \cap Y_n = \{x_{in}^1, x_{in}^2, \dots, x_{in}^{z_{in}}\}$ . Hence,

$$\begin{aligned}
 & \sum_{i=1}^m \sum_{j=1}^n \frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i \cap Y_j|}{|X_i|} \\
 &= \frac{1}{|U|} \sum_{i=1}^m \left( \log_2 \frac{|[x_{i1}^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} + \dots + \log_2 \frac{|[x_{i1}^{z_{i1}}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} \right) \\
 &+ \dots + \frac{1}{|U|} \sum_{i=1}^m \left( \log_2 \frac{|[x_{im}^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} + \dots + \log_2 \frac{|[x_{im}^{z_{im}}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} \right) \\
 &= \frac{1}{|U|} \left( \log_2 \frac{|[x_i^1]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^1]_{\mathcal{B}}|} + \dots \log_2 \frac{|[x_i^{k_i}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} + \dots + \log_2 \frac{|[x_i^{k_i}]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i^{k_i}]_{\mathcal{B}}|} \right) \\
 &= \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[x_i]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i]_{\mathcal{B}}|}
 \end{aligned} \tag{A3}$$

Therefore,

$$CH(\mathcal{D} | \mathcal{B}) = - \sum_{i=1}^m \sum_{j=1}^n \frac{|X_i \cap Y_j|}{|U|} \log_2 \frac{|X_i \cap Y_j|}{|X_i|} = - \frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[x_i]_{\mathcal{B} \cup \mathcal{D}}|}{|[x_i]_{\mathcal{B}}|} \tag{A4}$$

This completes the proof (i).

Theorem (ii) and (iii) are easy to know by the relationship of  $H(D; B) = H(D) - H(D | B)$  and  $H(D \cup B) = H(D) + H(B) - H(D | B)$ ; hence, they are omitted here.  $\square$

**References**

1. Ahmed, I.M.; Kashmoola, M.Y. CCF Based System Framework In Federated Learning Against Data Poisoning Attacks. *J. Appl. Sci. Eng.* **2022**, *26*, 973–981.
2. Lin, H.; Liu, G.N.; Wu, J.J.; Zuo, Y.; Wan, X.; Li, H. Fraud detection in dynamic interaction network. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1936–1950.
3. Shehnepoor, S.; Salehi, M.; Farahbakhsh, R. NetSpam: A network-based spam detection framework for reviews in online social media. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 1585–1595.
4. Dal-Pozzolo, A.; Caelen, O.; Le-Borgne, Y. A. Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Syst. Appl.* **2014**, *41*, 4915–4928.
5. Repousis, S.; Lois, P.; Veli, V. An investigation of the fraud risk and fraud scheme methods in Greek commercial banks. *J. Money Laund. Control.* **2019**, *22*, 53–61.
6. Tsang, S.; Koh, Y.S.; Dobbie, G.; Alam, S. SPAN: Finding collaborative frauds in online auctions. *Knowl. Based Syst.* **2014**, *71*, 389–408.

7. Pourhabibi, T.; Ong, K.; Kam, B.H. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decis. Support Syst.* **2020**, *133*, 113303.
8. Zhao, Q.; Chen, K.; Li, T. Detecting telecommunication fraud by understanding the contents of a call. *Cybersecurity* **2018**, *1*, 1–12.
9. Yang, J.; Yang, T.; Shi, C. Research on fault identification method based on multi-resolution permutation entropy and ABC-SVM. *J. Appl. Sci. Eng.* **2021**, *25*, 733–742.
10. Jurgovsky, J.; Granitzer, M.; Ziegler, K. Sequence classification for credit-card fraud detection. *Expert Syst. Appl.* **2018**, *100*, 234–245.
11. Wang, X.W.; Yin, S.L.; Li, H. A Network Intrusion Detection Method Based on Deep Multi-scale Convolutional Neural Network. *Int. J. Wireless Inf. Netw.* **2020**, *27*, 503–517.
12. Fiore, U.; De-Santis, A.; Perla, F. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Inf. Sci.* **2019**, *479*, 448–455.
13. Barnett, V.; Lewis, T. *Outliers in Statistical Data*; John Wiley and Sons: Hoboken, NJ, USA, 1994.
14. Wang, Y.; Li, Y. Outlier detection based on weighted neighbourhood information network for mixed-valued datasets. *Inf. Sci.* **2021**, *564*, 396–415.
15. M.M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander, LOF: Identifying Density-Based Local Outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000; pp. 93–104.
16. Ali, D.; Omer K. Efficient density and cluster based incremental outlier detection in data streams. *Inf. Sci.* **2022**, *607*, 901–920.
17. Li, Z.; Qu, L.; Zhang, G. Attribute selection for heterogeneous data based on information entropy. *Int. J. Gen. Syst.* **2021**, *50*, 548–566.
18. Salehi, F.; Keyvanpour, M.R.; Sharifi, A. SMKFC-ER: Semi-supervised multiple kernel fuzzy clustering based on entropy and relative entropy. *Inf. Sci.* **2021**, *547*, 667–688.
19. Wang, J.; Wei, J.; Yang, Z. Feature selection by maximizing independent classification information. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 828–841.
20. Thuy, N.N.; Wongthanasu, S. On reduction of attributes in inconsistent decision tables based on information entropies and stripped quotient sets. *Expert Syst. Appl.* **2019**, *137*, 308–323.
21. Patrick, G.C.; Cheng, G.; Jerzy, W.G.; Teresa, M. Characteristic sets and generalized maximal consistent blocks in mining incomplete data. *Inf. Sci.* **2018**, *453*, 66–79.
22. Liu, G.N.; Guo, J.; Zuo, Y.; Wu, J.J.; Guo R.Y. Fraud detection via behavioral sequence embedding. *Knowl. Inf. Syst.* **2020**, *62*, 2685–2708.
23. Hu, X.X.; Chen, H.C.; Liu, S.X.; Jiang, H.C.; Chu, G.H.; Li, R. BTG: A Bridge to Graph machine learning in telecommunications fraud detection. *Fut Gen. Comp. Syst.* **2022**, *137*, 274–287.
24. Emmanuel, O.; Rose, O.A.; Mohammed, U.A.; Ezekiel, R.A.; Bashir, T.; Salihu, S. Detecting Telecoms Fraud in a Cloud-Base Environment by Analyzing the Content of a Phone Conversation. *Asian J. Res. Comp. Sci.* **2022**, *4*, 115–131.
25. Viktoras, C.; Andrej, B.; Rima, K.; Olegas, V. Outlier Analysis for Telecom Fraud Detection. *Dig. Bus. Int. Syst.* **2022**, *1598*, 219–231.
26. Mollaoğlu, A.; Baltaoğlu, G.; Çakır, E.; Aktas, M.S. Fraud Detection on Streaming Customer Behavior Data with Unsupervised Learning Methods. In Proceedings of the 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia, 12–13 June 2021.
27. Zhong, Y.; Zhang, X.Y.; Shan, F. Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures. *Expert Syst. Appl.* **2018**, *112*, 243–257.
28. Qian, Y.; Liang, J.; Wei-zhi, Z.W. Information granularity in fuzzy binary GrC model. *IEEE Trans. Fuzzy Syst.* **2010**, *19*, 253–264.
29. Feng, Q.R.; Zhou, Y. Soft discernibility matrix and its applications in decision making. *Appl. Soft Comp.* **2014**, *24*, 749–756.
30. Shu, W.; Qian, W. An incremental approach to attribute reduction from dynamic incomplete decision systems in rough set theory. *Data Knowl. Eng.* **2015**, *100*, 116–132.
31. Sun, Y.; Mi, J.; Chen, J. A new fuzzy multi-attribute group decision-making method with generalized maximal consistent block and its application in emergency management. *Knowl. Based Syst.* **2021**, *215*, 106594.
32. Zhao, X.; Zhang, J.; Qin, X. LOMA: A local outlier mining algorithm based on attribute relevance analysis. *Expert Syst. Appl.* **2017**, *84*, 272–280.
33. Liang, B.H.; Liu, Y.; Shen, C.Y. Attribute Reduction Algorithm Based on Indistinguishable Degree. In Proceedings of the 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 20–22 April 2018.
34. Luo, C.; Li, T.; Huang, Y. Updating three-way decisions in incomplete multi-scale information systems. *Inf. Sci.* **2019**, *476*, 274–289.
35. Du, W.S.; Hu, B.Q. Attribute reduction in ordered decision tables via evidence theory. *Inf. Sci.* **2016**, *364*, 91–110.
36. Lang, G.; Cai, M.; Fujita, H. Related families-based attribute reduction of dynamic covering decision information systems. *Knowl. Based Syst.* **2018**, *162*, 161–173.
37. Liang, J.; Shi, Z.; Li, D. Information entropy, rough entropy and knowledge granulation in incomplete information systems. *Int. J. Gen. Syst.* **2006**, *35*, 641–654.
38. Liu, X. Research on Uncertainty Measurement and Attribute Reduction in Generalized Fuzzy Information Systems. Ph.D. Thesis, Hunan Normal University, Changsha, China, 2022.
39. Dai, J.; Liu, Q. Semi-supervised attribute reduction for interval data based on misclassification cost. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 1739–1750.

40. Sun, L.; Xu, J.C.; Yun, T. Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. *Knowl. Based Syst.* **2012**, *36*, 206–216.
41. Gao, C.; Zhou, J.; Miao, D.Q.; Yue, X. D.; Wan, J. Granular-conditional-entropy-based attribute reduction for partially labeled data with proxy labels. *Inf. Sci.* **2021**, *580*, 111–128.
42. Wang, Y.B.; Chen, X.J.; Dong, K. Attribute reduction via local conditional entropy. *Int. J. Mach. Learn. Cyb.* **2019**, *10*, 3619–3634.
43. Qu, L. D.; He, J. L.; Zhang, G. Q.; Xie, N. X. Entropy measure for a fuzzy relation and its application in attribute reduction for heterogeneous data. *Appl. Soft Comp.* **2022**, *118*, 108455.
44. Yang, H.; Moody, J. Data visualization and feature selection: New algorithms for nongaussian data. *Adv. Neural Inf. Process. Syst.* **1999**, *12*, 687–693.
45. Jakulin, A. Machine Learning Based on Attribute Interactions. Ph.D. Thesis, University of Ljubljana, Ljubljana, Slovenia, 12 May 2006.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.