mathematics

MDPI

Article

Knowledge Granularity Attribute Reduction Algorithm for Incomplete Systems in a Clustering Context

Baohua Liang [1,2,3], Erli Jin [3], Liangfen Wei [3] and Rongyao Hu [4,*]

[1] Guangxi Key Lab of Multi-Source Information Mining & Security, Guangxi Normal University, Guilin 541004, China; 054029@chu.edu.cn
[2] School of Computer Science and Engineering, Guangxi Normal University, Guilin 541004, China
[3] School of Computer and Artificial Intelligence, Chaohu University, Hefei 238000, China; 005044@chu.edu.cn (E.J.); 054073@chu.edu.cn (L.W.)
[4] CBICA, University of Pennsylvania, Philadelphia, PA 19104, USA
* Correspondence: rongyao.hu@pennmedicine.upenn.edu

Abstract: The phenomenon of missing data can be seen everywhere in reality. Most typical attribute reduction models are only suitable for complete systems. But for incomplete systems, we cannot obtain the effective reduction rules. Even if there are a few reduction approaches, the classification accuracy of their reduction sets still needs to be improved. In order to overcome these shortcomings, this paper first defines the similarities of intra-cluster objects and inter-cluster objects based on the tolerance principle and the mechanism of knowledge granularity. Secondly, attributes are selected on the principle that the similarity of inter-cluster objects is small and the similarity of intra-cluster objects is large, and then the knowledge granularity attribute model is proposed under the background of clustering; then, the IKAR algorithm program is designed. Finally, a series of comparative experiments about reduction size, running time, and classification accuracy are conducted with twelve UCI datasets to evaluate the performance of IKAR algorithms; then, the stability of the Friedman test and Bonferroni–Dunn tests are conducted. The experimental results indicate that the proposed algorithms are efficient and feasible.

Keywords: attribute reduction; knowledge granularity; clustering; similarity

MSC: 74H10

Citation: Liang, B.; Jin, E.; Wei, L.; Hu, R. Knowledge Granularity Attribute Reduction Algorithm for Incomplete Systems in a Clustering Context. *Mathematics* **2024**, *12*, 333. https://doi.org/10.3390/math12020333

Academic Editor: Junzo Watada

Received: 22 December 2023
Revised: 16 January 2024
Accepted: 18 January 2024
Published: 19 January 2024

1. Introduction

Rough set theory (RST) [1], initiated by Pawlak, is an effectively mathematical tool to deal with imprecise, fuzzy, and incomplete data. RST has been successfully applied in machine learning [2–4], knowledge discovery [5,6], expert system [7], disease diagnostics [8–10], decision support [11–13], and other areas [14,15]. Attribute reduction is one of the research hotspots in RST. As an important technology in the process of data preprocessing, attribute reduction has captured researchers' attention in big data and knowledge discovery [16–18]. The main objective of attribute reduction is to remove some irrelevant or non-important attributes while keeping the original distinguishing ability unchanged. In this way, the effect of data dimension reduction can be achieved, and a lot of time and space resources can be saved for the process of knowledge discovery and rule extraction.

With the rapid development of network and information technology, we have gradually entered the era of big data. The datasets have the characteristics of large volume, rapid change, and diverse data forms [19]. At the same time, due to the influence of the collection method and environment during the data collection process, there are a large number of missing data or wrong data in the datasets. The existence of these disturbed data will seriously affect the decision making and judgement of big data, and even mislead decision makers. After a long period of unremitting endeavor, scholars have achieved outstanding

results in attribute reduction [20–26]. For example, Dai [20] proposed a semi-supervised attribute reduction based on attribute indiscernibility. Cao [21] put forward a three-way approximate reduction approach by using information-theoretic measure. Yang [22] presented a novel incremental attribute reduction method via quantitative dominance-based neighborhood self-information. Lin [16] et al. developed a feature selection way by using neighborhood multi-granulation fusion. In a variable precision rough set model, Yu [24] et al. raised a novel attribute reduction based on local attribute significance. In the literature [27], Devi proposed a new dimension reduce technology by considering the picture of fuzzy soft matrices in the decision-making process. Wen [28] et al. raised an unsupervised attribute reduction algorithm for mixed data based on fuzzy optimal approximation set. These above-mentioned classic reduction models are suitable only for complete systems.

As with most of the datasets in various real-world applications, the classical rough set model defined with equivalence relation leads to the limitation in handling data in incomplete systems. The important attributes cannot be selected correctly, which leads to a decrease in the classification accuracy of the reduction set. In order to reduce the incomplete system, Zhang and Chen proposed a lambda-reduction approach based on the similarity degree respect to a conditional attribute subset for incomplete set-valued information systems [25]. For incomplete interval-valued information systems, Li [29] proposed the concept of similarity degree and tolerance relation between two information values of a given attribute. Then, three reduction algorithms based on theta-discernibility matrix, theta-information entropy, and theta-significance were designed. Liu introduced a new attribute reduction approach by using conditional entropy based on the fuzzy alpha-similarity relation [30]. Subsequently, Dai [31] proposed interval-valued fuzzy min–max similarity relations and designed two attribute reduction algorithms based on interval-valued fuzzy discernibility pairs model. Song [32] put forward the similarity degree between information values on each attribute and an attribute reduction method was designed by using information granulation and information entropy. Zhou presented a heuristic attribute reduction algorithm with a binary similarity matrix and attribute significance as heuristic knowledge under incomplete information systems [33]. Zhang [34] presented a novel approach for knowledge reduction by using the discernibility techniques in multi-granulation rough set model. He and Qu [35] put forward the fuzzy-rough iterative computation model based on symmetry relations for an incomplete categorical decision information system. Srirekha et al. proposed an attribute reduction in SE-ISI concept lattice based on the concept of object ranking for incomplete information systems [36]. Cornelis et al. put forward a generalized model of attribute reduction using fuzzy tolerance relation within the context of fuzzy rough set theory [37]. Liu applied the concept of accurate reduction and reduced invariant matrix for reducing attribute under information systems [38]. To reduce unnecessary tolerance classes for the original cover, Nguyen [39] introduced a new concept of stripped neighborhood covers and proposed an efficient heuristic algorithm in mixed and incomplete decision tables.

Although the above reduction algorithms can effectively reduce incomplete information systems, the classification accuracy of the reduction set is not ideal. The main reason is that only the importance of attributes is considered when selecting attributes, and the impact of attributes on classification is not considered. Usually, people take the best result of clustering as a reference standard for classification work and classify similar samples into the same cluster. In order to solve the problems mentioned above, this paper proposes an attribute reduction method from the perspective of clustering.

At present, the studies of attribute reduction on using the clustering idea to construct a feature selection model are relatively infrequent. In order to avoid or reduce the loss of some original information after discretizing continuous values, Zhang [40] proposed a feature selection method based on fuzzy clustering, but this method has no reliable theoretical support. Jia [41] proposed a spectral clustering method based on neighborhood information entropy feature selection, which uses a feature selection method to remove redundant features before clustering. In order to take the classification effect of the dataset

into consideration when selecting features, Zhao proposed a fuzzy C-Means clustering fuzzy rough feature selection method [42], which can improve the classification accuracy of the reduction set to a certain degree, but the effect is not obvious. Jia proposed a similarity attribute reduction in the context of clustering in the literature [43], which can greatly improve the classification accuracy of the reduction set but needs to continuously adjust the parameters to achieve the best classification effect. Such a feature set has certain random limitations and increases the time consumption of the system. Therefore, it is necessary to design a stable model with high classification accuracy for data preprocessing.

Although these existing approaches can effectively reduce incomplete systems, they only consider the importance of the attributes themselves, and do not consider the correlation between attributes. The influence of conditional attributes on decision classification is not considered. In order to improve the classification accuracy of a reduction set, we apply the idea of clustering.

Based on the principle that the similarity of samples within a cluster is as large as possible and the similarity of samples between clusters is as small as possible, an attribute reduction algorithm for an incomplete system is designed under the background of clustering. First, according to the principle of tolerance and the theory of knowledge granularity, we define the similarity of intra-cluster and inter-cluster for an incomplete system. Secondly, a formula for calculating the similarity of intra-cluster and inter-cluster objects is designed. After normalizing the two similarities, we define the similarity of objects. Then, according to the corresponding similarity mechanism, a new attribute reduction algorithm for an incomplete system is proposed. Finally, a series of experiments have verified that the proposed algorithm in this paper is significantly better than other similar algorithms in terms of running time, accuracy, and the stability of algorithm was analyzed by using Friedman test and Bonferroni–Dunn test in statistics.

The contribution of this paper is embodied in the following four aspects:

(1) A tolerance class calculation in incomplete information systems is proposed and applied to knowledge granularity calculation.
(2) Knowledge granules are used as a measure of sample similarity to measure the similarity of inter-cluster samples and intra-cluster samples.
(3) A knowledge granularity reduction algorithm based on clustering context is designed in incomplete information systems.
(4) Lots of experiments are conducted to verify the validity of the algorithm proposed in this paper, and the stability of the algorithm is verified by mathematical statistics.

The other parts of this paper are constructed as follows. The principle of tolerance and related concepts of knowledge granularity are recalled in Section 2. In Section 3, we propose a similarity measure of intra-cluster and inter-cluster objects and discuss the reduction mechanism according to the clustering background for missing dataset. We normalize the similarity of inter-cluster and intra-cluster, and then design the corresponding reduction model in Section 4. In Section 5, a series of experiments are conducted and the performance of the algorithm is evaluated from the reduction size, running time, classification accuracy, and stability. Then, the feasibility and effectiveness of the algorithm are verified. Finally, the advantages and disadvantages of the algorithm proposed in this paper are concluded and unfolded in the future work.

2. Preliminaries

In this section, we review some basic concepts in rough set theory, the definitions of tolerance class, knowledge granularity, clustering metrics, and the significance of attribute for incomplete decision systems.

2.1. Basic Concept of RST

A decision information system is a quadruple $DS = (U, A, V, f)$, where U is a nonempty finite set of objects and A is a finite nonempty attribute sets; if $A = C \cup D$, where C is the conditional attribute sets and D is the decision attribute set; then, V is the union of

attribute domains, $V = \cup_{a \in A} V_a$, V_a is the value set of attribute a, called the domain of a; $f : U \times A \to V$ is an information function with $f(x, a) = V_a$ for each $a \in A$ and $x \in U$. For every attribute subset $B \subseteq C$, a indiscernibility relation is defined as follows:

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

By the relation $IND(B)$, we can obtain the partition of U denoted by $U/IND(B)$ or U/B. If $B \subseteq A \wedge X \subseteq U$, the upper approximation is denoted as

$$\overline{B}(X) = \{x \in U | [x]_P \cap X \neq \varnothing\}. \tag{2}$$

The lower approximation of X can be denoted as

$$\underline{B}(X) = \{x \in U | [x]_P \subset X\} \tag{3}$$

where the objects in $\overline{B}(X)$ may belong to X, while the objects in $\underline{B}(X)$ must belong to X.

Definition 1. *In a decision system $DS = (U, C, D, V, f)$, if $\exists a \in C$ and $\exists x \in U$ that $f(x, a) = *$, then we call the decision system an incomplete system (IDS). In an incomplete decision system, if $P \subseteq C \cup D$ the tolerance relation is as follows:*

$$\mathrm{T(P)} = \{(\mathrm{x}, \mathrm{y}) \in \mathrm{U} \times \mathrm{U} | \forall \mathrm{a} \in \mathrm{P}, \mathrm{f}(\mathrm{x}, \mathrm{a}) = \mathrm{f}(\mathrm{y}, \mathrm{a}) \vee \mathrm{f}(\mathrm{x}, \mathrm{a}) = * \vee \mathrm{f}(\mathrm{y}, \mathrm{a}) = *\} \tag{4}$$

*where * represents missing value. $T(P)$ is symmetric and reflexive, but not transitive.*

Definition 2. *Given an incomplete decision system $IDS = (U, C \cup D, V, f)$, $\forall P \subseteq C \cup D$, $T(P) = \underset{a \in B}{\cap} T(a)$, and $T_P(o)$ is the tolerance class determined by o with respect to P, which is defined as follows:*

$$T_P(o) = \{y \in U | (o, y) \in T(P)\}. \tag{5}$$

Let $U/T(P)$ denote the family set of $T_P(o)$, which is the classification included by P. If $P \subseteq B$, then $T_B(o) \subseteq T_P(o)$, which is the monotonicity of the tolerance class.

2.2. Basic Concept of Knowledge Granularity

Definition 3. *Suppose $IDS = (U, C \cup D, V, f)$ is an incomplete decision system, $\forall P \subseteq C$. $T_P(o_i)$ is the tolerance class of object o_i with respect to P. The knowledge granularity of P on U is defined as follows:*

$$GK_U(P) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |T_P(o_i)| \tag{6}$$

where $|U|$ represents the number of objects in dataset U. Since the reflexivity and symmetry of $T_P(o_i)$, there are lot of repeated calculations when calculating $\sum_{i=1}^{|U|} |T_P(o_i)|$. In order to reduce the amount of calculation, we propose Definition 4 as follows:

Definition 4. *Suppose $IDS = (U, C \cup D, V, f)$ is an incomplete decision system, $\forall P \subseteq C$. $CT_P(o_i)$ is the simplified tolerance relation of object o_i with respect to P, which is defined as follows:*

$$\begin{aligned} CT_P(o_i) \quad &= \{o_j | \forall a \in P, (f(a, o_i) = f(a, o_j) \ \vee \ (\ f(a, o_i) \\ &= * \wedge f(a, o_j) = *) \wedge (i < j)) \ \vee (f(a, o_i) = * \wedge f(a, o_j) \neq *)\} \end{aligned}. \tag{7}$$

Definition 5. *Given an incomplete decision system IDS=$(U, C \cup D, V, f)$, $\forall P \subseteq C$ and $o \in U$, the simplified tolerance class of object o with respect to attribute P is defined as follows:*

$$CT_P(o) = \{y | (o, y) \in CT(P)\}. \tag{8}$$

We can delete the symmetric element pair and reflexive element pair from Definition 3 and obtain Definition 5, so that Definition 5 does not have the characteristics of symmetry and reflexivity.

Definition 6. *Given an incomplete decision system IDS* $=(U, C \cup D, V, f), \forall P \subseteq C$, $o \in U$,$CT_P(o)$, *is the simplified tolerance class of object o with respect to attribute P. The equal knowledge granularity of P on U is defined as follows:*

$$EGK_U(P) = \frac{1}{|U|^2} \sum_{o \in U} |CT_P(o)|. \tag{9}$$

Theorem 1. *Given an incomplete decision system IDS* $=(U, C \cup D, V, f), \forall P \subseteq C$. *Let* $U/P = \{X_1, X_2, \cdots, X_l\}, X_i \subseteq U/P$, $|X_i| = n_i$ *where* $1 \prec i \prec l$. *All objects in subdivision* X_l *are missing a value on attribute P,* $|X_l| = n_*$. *For the convenience of the following description, we mark* X_l *as* X_*. *Objects with all non-missing values on attribute P are marked with* $\overline{X_*}$. $EGK_U(P)$ *represents the equal knowledge granularity of P on U, we have:*

$$EGK_U(P) = \frac{1}{|U|^2} \left(\sum_{i=1}^{l} C_{n_i}^2 + |X_*||\overline{X_*}| \right) \tag{10}$$

where $C_n^2 = \frac{n(n-1)}{2}$.

Proof. Suppose that $\forall o \in X_i$ and $\forall o_* \in X_*$. According to Definition 4, we can obtain $\sum_{o \in X_i} |CT_P(o)| = \frac{n_i(n_i-1)}{2} + n_i n_*$ and $\sum_{o_* \in X_*} |CT_P(o_*)| = \frac{n_*(n_*-1)}{2}$. Suppose $\forall o \in U$, according to Definition 5, then we can obtain $EGK_U(P) = \frac{1}{|U|^2} \sum_{o \in U} |CT_P(o)| = \frac{1}{|U|^2} \left(\sum_{i=1}^{l-1} \sum_{o \in X_i} |CT_P(o)| + \sum_{o_* \in X_*} |CT_P(o_*)| \right) = \frac{1}{|U|^2} \cdot \left(\sum_{i=1}^{l-1} \frac{n_i(n_i-1)}{2} + n_* \sum_{i=1}^{l-1} n_i + \frac{n_*(n_*-1)}{2} \right) = \frac{1}{|U|^2} \left(\sum_{i=1}^{l} \frac{n_i(n_i-1)}{2} + n_*(|U| - n_*) \right)$. Since $C_{n_i}^2 = \frac{n_i(n_i-1)}{2}$, $C_{n_*}^2 = \frac{n_*(n_*-1)}{2}$ and $|\overline{X_*}| = |U| - n_*$, we can obtain $EGK_U(P) == \frac{1}{|U|^2} \left(\sum_{i=1}^{l} C_{n_i}^2 + |X_*||\overline{X_*}| \right)$. □

Property 1. *Given an incomplete decision system IDS=* $(U, C \cup D, V, f)$, $\forall P \subseteq C$. *If the knowledge granularity of P is* $GK_U(P)$ *on U and the equal knowledge granularity of P is* $EGK_U(P)$, *then we have:*

$$EGK_U(P) = (GK_U(P) - \frac{1}{|U|})/2. \tag{11}$$

Proof. Let $U/P = \{X_1, X_2, \cdots, X_{l-1}, X_*\}$, $|X_i| = n_i, |X_*| = n_*$ where X_i is the i-th subdivision of U/P and X_* stands for the subdivision of missing values on attribute P, $|X_i|$ stands for the number of object in subdivision X_i. We can obtain $|U| - \sum_{i=1}^{l-1} |X_i| = |X_*|$. According to Definition 2, suppose that $\forall o \in X_i$ and $T_P(o) = \{x | x \in X_i \vee x \in X_*\}$, we can obtain $|T_P(o)| = n_i + n_*$. Since the $|T_P(o)|$ value of each object in X_i is $n_i + n_*$, we can obtain $\sum_{o \in X_i} |T_P(o)| = n_i(n_i + n_*)$. In the same way, we can obtain $\sum_{o \in X_*} |T_P(x)| = n_*|U|$. According to Definition 3, we can obtain $GK_U(P) = \frac{1}{|U|^2} (\sum_{i=1}^{l-1} \sum_{o \in X_i} |T_P(o)| + \sum_{o \in X_*} |T_P(o)|) = \frac{1}{|U|^2} \cdot \left[\sum_{i=1}^{l-1} n_i(n_i + n_*) + n_*|U| \right]$, then $GK_U(P) - \frac{1}{|U|} = \frac{1}{|U|^2} (\sum_{i=1}^{l-1} (n_i^2 - n_i) + \sum_{i=1}^{l-1} n_i + n_* \sum_{i=1}^{l-1} n_i + n_*|U| - |U|) = \frac{1}{|U|^2} (\sum_{i=1}^{l-1} (n_i^2 - n_i) - n_*^2 - n_* + 2n_* |U|) = \frac{2}{|U|^2} \left(\sum_{i=1}^{l-1} \frac{(n_i^2 - n_i)}{2} + \frac{n_*^2 - n_*}{2} + n_*(|U| - n_*) \right) =$

$2EGK_U(P)$. Let $|X_l| = |X_*|$, we can obtain $GK_U(P) - \frac{1}{|U|} = \frac{2}{|U|^2}\left(\sum_{i=1}^{l}\frac{(n_i^2-n_i)}{2} + n_*(|U|-n_*)\right) = 2EGK_U(P)$. □

Due to the time complexity of calculating $T_P(o)$ is $|U|$, we know that the time complexity of calculating $GK_U(P)$ is $|U|^2$. However, the time complexity of $CT_P(o)$ is $|U/P|$, the time complexity of calculating $EGK_U(P)$ is $|U/P|^2$, and $|U/P|^2 \ll |U|^2$. In addition, in the process of calculating $EGK_U(P)$, the sub-division with a cardinality of 1 is constantly pruned, which further speeds up the calculation. Therefore, the time of calculating $EGK_U(P)$ is less than $GK_U(P)$ for the same data set.

Example 1. *Example of computing equivalent knowledgegranularity. Let* $IDS = (U, C \cup D, V, f)$, $U = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8, o_9\}$, $C = \{a, b, c, e\}$, $D = \{d\}$. *The detailed data are shown in Table 1. Let* $P = \{a, b\}$, *we can obtain* $f(o_6, P) = *$, $f(o_9, P) = *$. *We use the following two methods to calculate* $EGK_U(P)$.

(1) *According to Definition 5, we obtain that* $CT_P(o_1) = \{o_2, o_6, o_9\}, CT_P(o_2) = \{o_6, o_9\}$, $CT_P(o_3) = \{o_4, o_5, o_6, o_9\}, CT_P(o_4) = \{o_5, o_6, o_9\}, CT_P(o_5) = \{o_6, o_9\}, CT_P(o_6) = \{o_9\}, CT_P(o_7) = \{o_6, o_8, o_9\}, CT_P(o_8) = \{o_6, o_9\}, CT_P(o_9) = \{\varnothing\}$. *According to Definition 6, we can obtain* $EGK_U(P) = \frac{1}{|U|^2}\cdot\sum_{i=1}^{9}|CT_P(o_i)| \frac{1}{9^2}(3+2+4+3+2+3+2+1+0) = \frac{20}{81}$.

(2) *Since* $U/P = \{\{o_1, o_2\}, \{o_3, o_4, o_5\}, \{o_7, o_8\}, \{o_6, o_9\}\}$, *let* $X_1 = \{o_1, o_2\}, X_2 = \{o_3, o_4, o_5\}$, $X_3 = \{o_7, o_8\}$, $X_* = \{o_6, o_9\}$, *then* $|X_*| = 2, |\overline{X_*}| = |U| - 2 = 7$. *According to Theorem 1, we can obtain* $EGK_U(P) = \frac{1}{9^2}\cdot[C_2^2 + C_3^2 + C_2^2 + C_2^2 + 2\cdot(9-2)] = \frac{20}{81}$.

Table 1. Incomplete information system.

U	a	b	c	e	d
o1	0	0	0	1	0
o2	0	0	1	*	1
o3	0	1	0	1	0
o4	0	1	*	0	1
o5	0	1	0	*	1
o6	*	*	1	1	1
o7	1	0	*	0	2
o8	1	0	1	0	2
o9	*	*	0	1	2

'*' represents the missing value about some attributes.

Although the above two methods achieve the same results, the calculation time is different. Since method 1 needs to scan the dataset multiple times, it consumes more time. However, method 2 only needs to scan the dataset one time and obtains each subdivision of U/P. According to the number of objects in each subdivision, we can acquire the combination value quickly, and then the value of equivalent knowledge granularity is calculated.

3. The Mechanism of Knowledge Granularity Attribute Reduction in the Background of Clustering

Most traditional attribute reduction models use equivalence class relation to compute the importance of conditional attributes. Although these methods can effectively deal with complete decision-making information systems, they cannot obtain correct reduction rules in incomplete ones. In order to deal with the loss of information effectively, this paper focuses on the reduction in incomplete decision systems.

The traditional reduction model does not consider the impact on the classification of the dataset when deleting redundant attributes. If there are inconsistent objects in the

dataset, the classification accuracy of the reduced set will be affected. In order to improve the data quality, this paper uses the idea of clustering. Clustering is to divide all objects in the dataset into different clusters according to a certain standard when the target category is unknown. Objects within a cluster are as similar as possible, and objects between clusters are as dissimilar as possible. Classification is to classify all objects in the dataset according to a certain nature and level when the object category is known. Good clustering results can be used as a reference standard for accurate classification. The desired results of clustering involve the objects of the same class being gathered in intra-clustering, otherwise they will be gathered in different inter-clustering. This paper studies the labeled data objects decision information system. Therefore, we use the results of the classification to guide the process of clustering the data objects. When the data objects are clustered, they follow the principle that the objects of intra-clustering are as close as possible and the objects of inter-clustering are as far away as possible. Next, we discuss how to measure the distance of intra-clustering and inter-clustering objects.

3.1. The Intra-Cluster Similarity for Incomplete Systems

Generally, there are two approaches for clustering calculations: distance and similarity. The closer the distance between two different objects is, the weaker their ability to distinguish is. On the contrary, the farther the distance is, the stronger the ability to distinguish is. In this paper, the similarity method is used to measure the distinguishing ability of objects. Since the knowledge granularity can measure the similarity between objects, the coarser the knowledge granularity is, the stronger the distinguishing ability is. The better the knowledge granularity is, the weaker the distinguishing ability is. Next, we discuss how to use knowledge granularity information to measure the similarity of objects in an incomplete system.

Definition 7 (The similarity of intra-cluster objects). *Given an incomplete decision system* $IDS = (U, C \cup D, V, f)$, $U/D = \{D_1, D_2, \cdots, D_n\}$. *For the sake of convenience, let* $U/D = \pi_D$, $D_i \subseteq \pi_D$, $P \subseteq C$. *Suppose the equivalence division relationship of* D_i *under attribute set P is* $\mathbb{R}_P = \{X_1, X_2, \cdots, X_m\}$; *then, the similarity of objects in the cluster of* D_i *about attribute P is defined as follows (where* $o \in D_i$*):*

$$SIntra_{D_i}(P) = EGK_{D_i}(P) = \frac{1}{|D_i|^2} \sum_{o \in D_i} |CT_P(o)|. \tag{12}$$

Definition 8 (The average similarity of intra-cluster objects). *Given an incomplete decision system* $IDS = (U, C \cup D, V, f)$, $U/D = \{D_1, D_2, \cdots, D_n\}$. $P \subseteq C$, $o \in D_i$. *The knowledge granularity similarity of intra-clustering objects for subdivision* D_i *with respect to attribute P is* $SIntra_{D_i}(P)$, *then the average intra-clustering similarity is defined as follows:*

$$ASIntra_{\pi_D}(P) = \frac{1}{n} \sum_{i=1}^{n} SIntra_{D_i}(P). \tag{13}$$

The desired effect of clustering is that the similarity of intra-clustering is high, and the similarity of inter-clustering is low.

Property 2. *Given an incomplete system* $IDS = (U, C \cup D, V, f), U/D = \pi_D, D_i \subseteq \pi_D, P, Q \subseteq C$. *If* $P \subseteq Q$, *and we have*

$$ASIntra_{\pi_D}(P) \geq ASIntra_{\pi_D}(Q). \tag{14}$$

Proof. Let $D_i/P = \{X_1, X_2, \cdots, X_k \cup X_{k+1}, X_{k+2}, \cdots, X_n, \ X_* \cup Y\}$, where X_*, Y represents the object sets with missing values on attribute set P. Since $P \subseteq Q$, we can obtain $D_i/Q \subseteq D_i/P$. This to say, each subdivision of D_i/Q is a subset of some subdivision of D_i/P. Let $D_i/Q = \{X_1, \ X_2, \cdots, X_k, X_{k+1}, X_{k+2}, \cdots, X_n, Y, X_*\}$. According to The-

orem 1 and Definition 7, we can obtain: $SIntra_{D_i}(P) = \frac{2}{|D_i|^2}(\sum_{j=1}^{k-1} C^2_{|X_j|} + C^2_{|X_k|+|X_{k+1}|} + \sum_{j=k+2}^{n} C^2_{|X_j|} + C^2_{|X_*|+|Y|} + (|X_*|+|Y|)(|D_i|-|X_*|-|Y|)) = \frac{2}{|D_i|^2}(\sum_{j=1}^{k-1} C^2_{|X_j|} + C^2_{|X_k|} + C^2_{|X_{k+1}|} + |X_k||X_{k+1}| + \sum_{j=k+2}^{n} C^2_{|X_j|} + C^2_{|X_*|} + C^2_{|Y|} + |X_*||Y| + (|X_*|+|Y|)(|D_i|-|X_*|-|Y|)) = \frac{2}{|D_i|^2}(\sum_{j=1}^{n} C^2_{|X_j|} + |X_k||X_{k+1}| + C^2_{|X_*|} + C^2_{|Y|} + |X_*||Y| + (|X_*|+|Y|)(|D_i|-|X_*|-|Y|))$. Since $SIntra_{D_i}(Q) = \frac{2}{|D_i|^2}(\sum_{j=1}^{n} C^2_{|X_j|} + C^2_{|X_*|} + C^2_{|Y|} + |X_*|(|D_i|-|X_*|))$, then $\frac{2}{|D_i|^2}(|X_k||X_{k+1}| + |Y||D_i| - |X_*||Y| - |Y|^2) = \frac{2}{|D_i|^2}(|X_k||X_{k+1}| + |Y|(|D_i|-|X_*|-|Y|))$. Since $|D_i| >= |X_*| + |Y|$, $|X_k||X_{k+1}| \geq 0$; therefore, the results of $SIntra_{D_i}(P) \geq SIntra_{D_i}(Q)$ and $\sum_{i=1}^{n} SIntra_{D_i}(P) \geq \sum_{i=1}^{n} SIntra_{D_i}(Q)$ are obtained. Above all, we can obtain $ASIntra_{\pi_D}(P) \geq ASIntra_{\pi_D}(Q)$. □

According to Property 2, we conclude that the intra-cluster similarity is monotonic when the conditional attribute set changes.

Example 2. *In Table 1, let $P = \{a, b\}$; we have $D_1 = \{o_1, o_3\}, D_2 = \{o_2, o_4, o_5, o_6\}, D_3 = \{o_7, o_8, o_9\}$,$SIntra_{D_1}(P) = 0$,$SIntra_{D_2}(P) = \frac{0+C_2^2+1\times 3}{4^2} = \frac{1}{4}$,$SIntra_{D_3}(P) = \frac{C_2^2+1\times 2}{3^2} = \frac{1}{3}$, $ASIntra_{\pi_D}(P) = \frac{1}{3}\sum_{i=1}^{3} SIntra_{D_i}(P) = \frac{7}{36}$.*

3.2. The Inter-Cluster Similarity for Incomplete Systems

Definition 9 (The inter-cluster similarity for incomplete systems). *Given $IDS = (U, C \cup D, V, f)$ is an incomplete decision system, let $\pi_D = \{D_1, D_2, \cdots, D_n\}$. Suppose $P \subseteq C$, $D_i, D_j \subset U/D$, then the inter-cluster similarity of D_i and D_j with respect to attribute set P for incomplete systems is defined as the following:*

$$SInter_{D_i,D_j}(P) = \frac{1}{(|D_i| + |D_j|)^2}\left(\sum_{o\in D_i\cup D_j} |CT_P(o)| - \sum_{o\in D_i} |CT_P(o)| - \sum_{o\in \cup D_j} |CT_P(o)|\right). \quad (15)$$

Assuming that D_i and D_j are objects of two different clusters, the inter-cluster similarity between D_i and D_j is calculated in two steps. The first step is to calculate the similarity after the two clusters are merged, and the second step is to remove the similarity information of the same cluster. The rest is the similarity information between objects in different clusters.

Property 3. *Given an incomplete decision system $IDS = (U, C \cup D, V, f)$, $\pi_D = \{D_1, D_2, \cdots, D_n\}$ $P \subseteq C$. Let $D_i/P = \{X_1, X_2, \cdots, X_k, X_{k+1}, \cdots, X_m, X_*\}, \overline{X_*} = D_i - X_*, D_j/P = \{Y_1, Y_2, \cdots, Y_k, Y_{k+1}, \cdots, Y_n, Y_*\}, \overline{Y_*} = D_j - Y_*, D_i \cup D_j/P = \{X_1 \cup Y_1, X_2 \cup Y_2, \cdots, X_k \cup Y_k, X_* \cup Y_*, X_{k+1}, \cdots, X_m, Y_{k+1}, \cdots, Y_n\}$, where $\forall o \in X_* \cup Y_*, f(o, P) = *$, '*' is the flag of missing value. We have:*

$$SInter_{D_i,D_j}(P) = \frac{1}{(|D_i| + |D_j|)^2}\left(\sum_{l=1}^{k} |X_l||Y_l| + +|X_*||Y_*| + |X_*|\overline{|Y_*|} + \overline{|X_*|}|Y_*|\right). \quad (16)$$

Proof. According to Definition 5 and Theorem 1, we can obtain $\sum_{o\in D_i\cup D_j}|CT_P(o)| = \sum_{l=1}^{k} C^2_{|X_l|+|Y_l|} + C^2_{|X_*|+|Y_*|} + \sum_{l=k+1}^{m} C^2_{|X_i|} + \sum_{l=k+1}^{n} C^2_{|Y_i|} + (|X_*|+|Y_*|)(|\overline{Y_*}|+|\overline{X_*}|)$, $\sum_{o\in D_i}|CT_P(o)| = \sum_{l=1}^{m} C^2_{|X_l|} + C^2_{|X_*|} + |X_*||\overline{X_*}|$, $\sum_{o\in D_j}|CT_P(o)| = \sum_{l=1}^{n} C^2_{|Y_l|} + C^2_{|Y_*|} + |Y_*||\overline{Y_*}|$. Since $\sum_{l=1}^{k} C^2_{|X_l|+|Y_l|} = \sum_{l=1}^{k}\left(\frac{|X_l|^2-|X_l|+|Y_l|^2-|Y_l|}{2} + |X_l||Y_l|\right) = \sum_{l=1}^{k} C^2_{|X_l|} + \sum_{l=1}^{k} C^2_{|Y_l|} + \sum_{l=1}^{k}|X_l||Y_l|$, according to Definition 9, we can conclude that $(|D_i|+|D_j|)^2\ SInter_{D_i,D_j}(P) = \sum_{o\in D_i\cup D_j}|CT_P(o)| - \sum_{o\in D_i}|CT_P(o)| - \sum_{o\in\cup D_j}|CT_P(o)| = \sum_{l=1}^{k}|X_l||Y_l| + |X_*||Y_*| + |X_*||\overline{Y_*}| + |\overline{X_*}||Y_*|$. Then, we have:

$$SInter_{D_i,D_j}(P) = \frac{1}{(|D_i|+|D_j|)^2}\left(\sum_{l=1}^{k}|X_l||Y_l| + |X_*||Y_*| + |X_*||\overline{Y_*}| + |\overline{X_*}||Y_*|\right).\square$$

Property 4. *Given an incomplete decision system $IDS = (U, C\cup D, V, f)$, $\pi_D = \{D_1, D_2, \cdots, D_n\}$, $P \subseteq Q \subseteq C, D_i, D_j \subset U/D$, then we have $SInter_{D_i,D_j}(P) \geq SInter_{D_i,D_j}(Q)$.*

Proof. Let $D_i/Q = \{X_1, X_2, \cdots, X_k, X_{k+1}, \cdots, X_m, X_\triangle,\ X_* - X_\triangle\}, \overline{X_* - X_\triangle} = D_i - X_* + X_\triangle, D_j/Q = \{Y_1, Y_2, \cdots,\ Y_k, Y_{k+1}, \cdots, Y_n, Y_\triangle, Y_* - Y_\triangle\}, \overline{Y_* - Y_\triangle} = D_j - Y_* + Y_\triangle$. Since $P \subseteq Q \subseteq C$, then D_i/Q is a refinement of D_i/P, D_j/Q is a refinement of D_j/P. Let $D_i/P = \{X_1, X_2, \cdots, X_k \cup X_{k+1}, \cdots,\ X_m, X_*\}, \overline{X_*} = D_i - X_*, D_j/P = \{Y_1, Y_2, \cdots, Y_k \cup Y_{k+1}, \cdots, Y_n, Y_*\}, \overline{Y_*} = D_j - Y_*$. Suppose $D_i \cup D_j/Q = \{\ X_1 \cup Y_1, X_2 \cup Y_2, \cdots, X_k \cup Y_k, X_{k+1} \cup Y_{k+1}, X_\Delta \cup Y_\Delta, (X_* - X_\Delta) \cup (Y_* - Y_\Delta), X_{k+2}, \cdots, X_m, Y_{k+2}, \cdots, Y_n\}, D_i \cup D_j/P = \{X_1 \cup Y_1, X_2 \cup Y_2, \cdots, X_k \cup Y_k, X_{k+1} \cup Y_{k+1}, X_* \cup Y_*, X_{k+2}, \cdots, X_m, Y_{k+2}, \cdots, Y_n\}$, then $SInter_{D_i,D_j}(P) = \sum_{l=1}^{k-1}|X_l||Y_l| + (|X_k|+|X_{k+1}|)(|Y_k|+|Y_{k+1}|) + |X_*||Y_*| + |X_*||\overline{Y_*}| + |\overline{X_*}||Y_*|$, $SInter_{D_i,D_j}(Q) = \sum_{l=1}^{k-1}|X_l||Y_l| + |X_k||Y_k| + |X_{k+1}||Y_{k+1}| + |X_\Delta||Y_\Delta| + |X - X_\Delta||Y - Y_\Delta| + |X_* - X_\Delta||\overline{Y_* - Y_\Delta}| + |\overline{X_* - X_\Delta}|\ |Y_* - Y_\Delta|$. $SInter_{D_i,D_j}(P) - SInter_{D_i,D_j}(Q) = |X_k||Y_{k+1}| + |X_{k+1}||Y_k| + |X_*||Y_*| + |X_*||\overline{Y_*}| + |\overline{X_*}|\ |Y_*| - |X_\Delta|\ |Y_\Delta| - |X - X_\Delta||Y - Y_\Delta| - |X_* - X_\Delta|\ |\overline{Y_* - Y_\Delta}| + |\overline{X_* - X_\Delta}||Y_* - Y_\Delta|$. Since $|\overline{X_*}| = |D_i| - |X_*|, |\overline{Y_*}| = |D_j| - |Y_*|, |\overline{X_* - X_\Delta}| = |D_i| - |X_*| + |X_\Delta|, |\overline{Y_* - Y_\Delta}| = |D_j| - |Y_*| + |Y_\Delta|, |X_k||Y_{k+1}| + |X_{k+1}||Y_k| \geq 0$, then $SInter_{D_i,D_j}(P) - SInter_{D_i,D_j}(Q) \geq |X_*||Y_*| + |X_*||\overline{Y_*}| + |\overline{X_*}||Y_*| - |X_\Delta||Y_\Delta| - |X - X_\Delta|\ |Y - Y_\Delta| - |X_* - X_\Delta||\overline{Y_* - Y_\Delta}| + |\overline{X_* - X_\Delta}||Y_* - Y_\Delta| = |X_\triangle| \cdot |D_j| + |Y_\Delta||D_i| - |X_*||Y_\Delta| - |X_\Delta||Y_*| = |X_\Delta|(|D_j| - |Y_*|) + |Y_\Delta| \cdot (|D_i| - |X_*|)$. Since $|D_j| \geq |Y_*|, |D_i| \geq |X_*|$, then we have $SInter_{D_i,D_j}(P) - SInter_{D_i,D_j}(Q) \geq 0$. $\square$

From Property 4, it can be concluded that the similarity of inter-clusters has monotonicity with the change in conditional attributes.

Definition 10 (The average similarity of inter-cluster objects). *Given an incomplete decision system $IDS = (U, C\cup D, V, f)$,$\pi_D = \{D_1, D_2, \cdots, D_n\}$, $P \subseteq C$. In n different clusters, the inter-cluster similarity is calculated between every two clusters and the time of comparisons is $\frac{1}{2}n(n-1)$, then the average knowledge granularity similarity of inter-cluster in dataset U respect of attribute set P is defined as the following:*

$$ASInter_{\pi_D}(P) = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} SInter_{D_i,D_j}(P). \tag{17}$$

Example 3. *In Table 1, with the same conditions as Example 2, let* $P = \{a, b\}$, *we have* $D_1 = \{o_1, o_3\}, D_2 = \{o_2, o_4, o_5, o_6\}, D_3 = \{o_7, o_8, o_9\}, SInter_{D_1,D_2}(P) = \frac{1\times1+1\times2+2\times1}{(2+4)^2} = \frac{5}{36}, SInter_{D_1,D_3}(P) = \frac{0+0+1\times2}{(2+3)^2} = \frac{2}{25}, SInter_{D_2,D_3}(P) = \frac{0+1\times2+3\times1}{(4+3)^2} = \frac{5}{49}, ASInter_{\pi_D}(P) = \frac{1}{3}\sum_{i=1}^{2}\sum_{j=i+1}^{3} SInter_{D_i,D_j}(P) = \frac{14153}{132300}$.

4. Attribute Reduction of Knowledge Granularity for Incomplete Systems

Traditional attribute reduction methods are mostly aimed at datasets with no missing data. Various datasets in reality are often incomplete due to various subjective or objective factors. Therefore, we researched information systems with missing data and propose corresponding algorithms to improve the data quality of incomplete system reduction sets.

This paper discusses how to design a reduction method with the idea of clustering. For an ideal clustering effect, the objects of inter-cluster should be far away, and the objects of intra-cluster should be close together. Here, similarity is used to measure the distance between two different objects. The higher the similarity is, the closer the objects are. Conversely, the lower the similarity is, the farther the objects are. Based on the above analysis, we designed a formula to measure the importance of attributes as the following:

$$SIM_R = SAIntra + \lambda\cdot(1 - ASInter) \tag{18}$$

where λ is the weight, and $1 - ASIntra$ is the dissimilarity of intra-cluster objects. We can set the importance of the intra-cluster similarity and inter-cluster similarity by using the size of the parameter λ. This method requires the adjustment of the parameters continuously, which consumes a lot of time. To this end, we first normalize $ASIntra$ and $ASInter$. Then, the two similarity calculations can be measured within a unified range, avoiding the parameter adjustment process.

4.1. Normalization of Inter-Cluster Similarity and Intra-Cluster Similarity

Given an incomplete decision system $IDS = (U, C \cup D, V, f), \pi_D = \{D_1, D_2, \cdots, D_n\}$, $P \subseteq C, D_i, D_j \subset U/D$. Since the number of elements in each sub-division may be different, then the value range of its equivalent knowledge granularity may also be different. To calculate the average similarity of all subdivisions, they must be calculated in the same domain. For the sake of generality, we normalize the inter-cluster similarity and intra-cluster similarity. According to Definition 7 and Theorem 1, we can obtain $SIntra_{D_i}(P) = EGK_{D_i}(P)$. When all data objects in the subdivision D_i are indistinguishable with respect to the attribute set P, $EGK_{D_i}(P) = \frac{1}{2}(1 - \frac{1}{|D_i|})$ takes the maximum value.

When all data objects in D_i can be distinguished from each other, $EGK_{D_i}(P) = 0$ takes the minimum value. So, the result of $EGK_{D_i}(P) \in \left[0, \frac{|D_i|-1}{2|D_i|}\right]$ is obtained. If the value of $EGK_{D_i}(P)$ is mapped to the range [0,1], the correction formula of $SIntra_{D_i}(P)$ is defined as

$$SIntra_{D_i}(P)' = EGK_{D_i}(P)/\left(\frac{|D_i|-1}{2|D_i|}\right) = EGK_{D_i}(P)\frac{2|D_i|}{|D_i|-1}. \tag{19}$$

The average similarity of intra-cluster objects is corrected as follows:

$$ASIntra_{\pi_D}(P) = \frac{1}{n}\sum_{i=1}^{n} SIntra_{D_i}(P)' \tag{20}$$

D_i and D_j are two object sets of different clusters. Suppose $D_i/P = \{X_1, X_2, \cdots, X_n, X_*\}$, $D_j/P = \{Y_1, Y_2, \cdots, Y_m, Y_*\}$. If $f(X_l, P) = f(Y_l, P)(1 \le l \le k)$, then $D_i \cup D_j/P =$

$\{X_1 \cup Y_1, X_2 \cup Y_2, \cdots, X_k \cup Y_k, \cdots, X_{k+1}, \cdots, X_n, Y_{k+1}, \cdots, Y_m, X_*, Y_*\}$. According to Property 3, the similarity of inter-cluster respect to D_i and D_j is denoted as

$$SInter_{D_i,D_j}(P) = \frac{1}{(|D_i| + |D_j|)^2}\left(\sum_{l=1}^{k}|X_l||Y_l| + |X_*|\overline{|Y_*|} + \overline{|X_*|}|Y_*|\right).$$

When $\cup_{i=1}^{k} X_i \cup Y_i = \varnothing, X_* = \varnothing, Y_* = \varnothing\}, SInter_{D_i,D_j}(P) = 0$, takes the minimum value. When all data objects in D_i and D_j are indistinguishable, $SInter_{D_i,D_j}(P) = \frac{|D_i||D_j|}{(|D_i|+|D_j|)^2}$ takes the maximum value. Then, we can obtain $SInter_{D_i,D_j}(P) \in \left[0, \frac{|D_i||D_j|}{(|D_i|+|D_j|)^2}\right]$. The normalized formula of $SInter_{D_i,D_j}(P)$ is as follows:

$$SInter_{D_i,D_j}(P)' = SInter_{D_i,D_j}(P)\frac{(|D_i| + |D_j|)^2}{|D_i||D_j|}. \tag{21}$$

The definition of the average similarity of inter-cluster objects is revised as follows:

$$ASInter_{\pi_D}(P) = \frac{n(n-1)}{2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} SInter_{D_i,D_j}(P)'. \tag{22}$$

After the similarities of inter-cluster and intra-cluster objects are normalized, $SIM_R = ASInter + \lambda\cdot(1 - ASIntra)$ is revised as follows:

$$SIM_R = AIntra_{\pi_D}(P) + 1 - AInter_{\pi_D}(P). \tag{23}$$

Since $ASInter_{\pi_D}(P)$ represents the similarity of intra-cluster objects, then the dissimilarity is $1-ASInter_{\pi_D}(P)$. If you use the formula of SIM_R to measure the effect of clustering, the larger the value of SIM_R is, the better the effect is.

4.2. The Knowledge Granularity Attribute Reduction Algorithm for Incomplete Systems (IKAR)

In Section 4.1, we discussed the similarities of inter-cluster and intra-cluster objects from the perspective of clustering, which provided a clear goal for the next step of attribute selection.

Definition 11 (Equal knowledge granularity attribute reduction). *Given an incomplete decision system $IDS = (U, C \cup D, V, f)$, an attribute subset $R \subseteq C$ is an equal know-ledge granularity attribute reduction if and only if:*

$$\begin{gathered}(1)R = \min_{P\subseteq C}\{SIM_P\} \\ (2)\forall R' \subset R, SIM_{R'} > SIM_R\end{gathered} \tag{24}$$

In Definition 11, condition (1) is the jointly sufficient condition that guarantees that the equal knowledge granularity value induced from the reduction is minimal, and condition (2) is the individual necessary condition that guarantees the reduction is minimal.

According to Definition 11, we can find a smaller reduction set with an ideal classification effect. Property 2 proves that when the attribute decreases, the similarity of intra-cluster increases monotonically. Property 4 proves that the similarity of inter-cluster also increases when decreasing the attribute, so that the dissimilarity will decrease. Obviously, the formula SIM_R cannot determine its monotonicity. To find the R when the SIM_R is the largest in the conditional attribute, which is a combinatorial optimization problem, and trying the methods one by one is not the best way to solve the problem. So, we use

a heuristic method to find the smallest reduced set. For any attribute $a \in C$, its inner significance is defined as follows:

$$Sig_C^a = SIM_C - SIM_{C-a} \tag{25}$$

The bigger the value of Sig_C^a is, the more important the attribute is.

In order to obtain the optimal reduction set quickly, we adopt the deletion strategy. Firstly, the importance Sig_C^a of attribute a is defined by the formula of SIM; then, sort the different Sig_C^a. Secondly, let R = C. The value of SIM_C is used as the initial condition, which ensures that the clustering effect after reduction is better than the raw dataset. Then, remove the unimportant attribute a from the remaining attributes $C - R$ and calculate the value of SIM_{R-a}. If $SIM_{R-a} \geq SIM_R$, delete attribute a and continue; otherwise, the algorithm terminates. The details of the IKAR algorithm are shown in Figure 1.

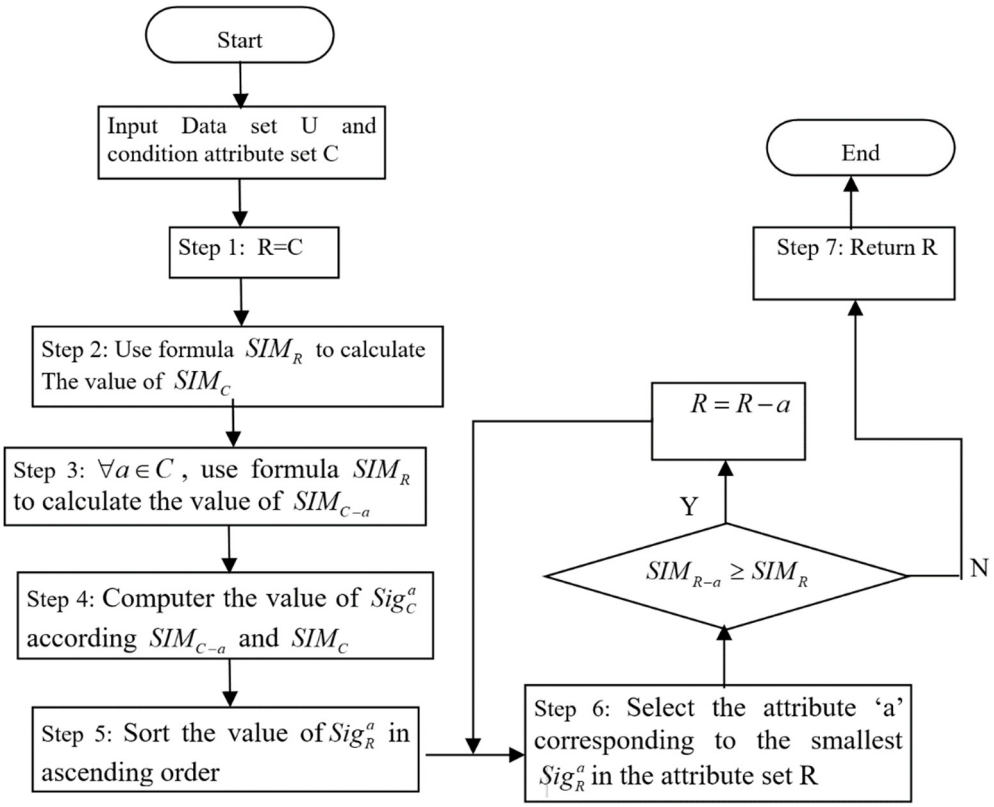

Figure 1. The execution process of the IKAR algorithm.

Example 4. *In Table 1, since* $D_1 = \{o_1, o_3\}, D_2 = \{o_2, o_4, o_5, o_6\}, D_3 = \{o_7, o_8, o_9\}$. *Let* $R = C$, *according to the definition of* SIM_R, *we can obtain* $SIM_R = \frac{72}{75}$, $SIM_{C-a} = \frac{71}{72}$, $SIM_{C-b} = \frac{100}{72}$, $SIM_{C-c} = \frac{60}{72}$ *and* $SIM_{C-e} = \frac{80}{72}$, *then* $Sig_C^a = \frac{4}{72}$, $Sig_C^b = -\frac{25}{72}$, $Sig_C^c = \frac{15}{72}$, $Sig_C^e = -\frac{5}{72}$. *Since* $SIM_{C-b} = \frac{100}{72} > SIM_R$, *we delete the attribute b from R and let* $R = R - b$, $SIM_R = \frac{100}{72}$. *In the same way, we obtain that* $SIM_{R-a} = \frac{94}{72}$, $SIM_{R-c} = \frac{92}{72}$, $SIM_{R-e} = \frac{108}{72}$, $Sig_R^a = \frac{6}{72}$, $Sig_R^c = \frac{8}{72}$ *and* $Sig_R^e = -\frac{8}{72}$. *Since* $SIM_{R-e} = \frac{108}{72} > SIM_R$, *we delete the attribute e from R and let* $R = R - e$. *We calculate the value of* SIM_{R-a}, SIM_{R-c}, *obtain the results of* $SIM_{R-a} = \frac{82}{72}$ *and* $SIM_{R-c} = \frac{100}{72}$. *Now, we have* $Sig_R^a = \frac{18}{72}$ *and* $Sig_R^c = \frac{8}{72}$. *If attribute c is deleted* $SIM_{R-c} = \frac{100}{72} < SIM_R$, *and the algorithm is terminated. We have* $R = \{a, c\}$.

4.3. Time Complexity Analysis

The time complexity of Step 1 is $O(1)$. In Step 2, we calculate the value of SIM_C which include $ASIntra_{\pi_D}(C)$ and $ASInter_{\pi_D}(C)$. The computational time complexity of intra-cluster $SIntra_{D_i}(C)$ similarity is $O(|C||D_i|)$, then the time complexity of calculat-

ing $ASIntra_{\pi_D}(C)$ is $O(|U||C|)$. Since the time complexity of calculating the similarity $SInter_{D_i,D_j}(C)$ about inter-cluster D_i and D_j is $O((|D_i| + |D_j|)|C|)$, then time complexity of $ASInter_{\pi_D}(C)$ is $O(\sum_{i=1}^{|U/D|-1}\sum_{j=i+1}^{|U/D|}(|D_i| + |D_j|)|C|) = O((|U/D| - 1)|U||C|)$. We can obtain the time complexity of Step 2 is $O(|U/D||U||C|)$. In Step 3, the consume time is $O(|U/D||U||C|)$. Since Step 4 utilizes the results of Step 3, the time complexity is $O(1)$. Step 5 is to sort the importance Sig_R^a of each attribute, and the time complexity is $O(\frac{1}{2}|C|^2)$. Since the time complexity of calculating Sig_R^a is $O(|U/D||U||R|)$, then the time complexity of Step 5 is $O(|U/D||U||R|)$+$O(\frac{1}{2}|C|^2)$. In Step 6, the time complexity of deleting a redundant attribute is $O(|U/D||U||R|)$, and Step 6 needs to be executed $|C| - |R|$ times, $R \subseteq C$, then the time complexity of Step 6 is $O(|U/D||U|(|C|^2 - |R|^2)$. In summary, the time complexity of IKAR is $O(|U/D||U|(|C|^2 - |R|^2) + \frac{1}{2}|C|^2)$.

5. Experiments Results Analysis

In order to evaluate the feasibility and effectiveness of the proposed algorithm in this paper, the complete dataset was preprocessed to obtain incomplete information systems, and many different attribute reduction algorithms were used for reduction. The reduction set obtained in the previous stage was classified and analyzed by multiple classifiers in the Weka Tool. It was compared with three other existing algorithms in terms of reduction set size, running time, accuracy, and algorithm stability. The specific experimental framework is shown in Figure 2.

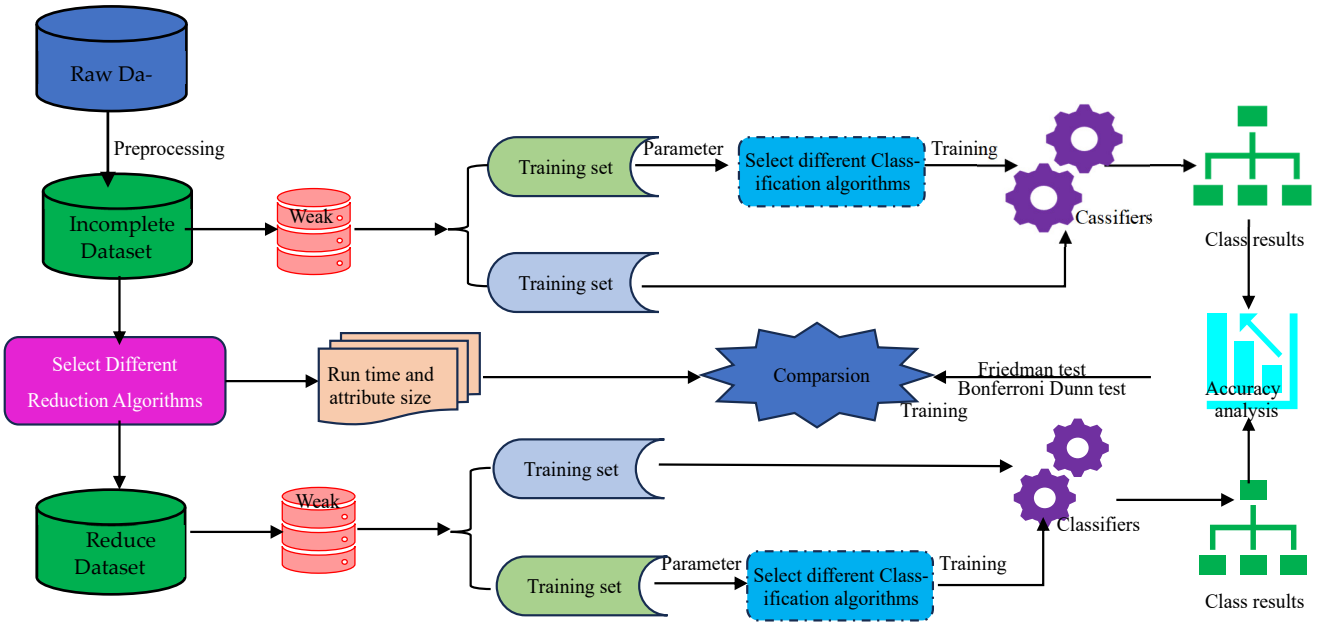

Figure 2. The framework chart of experiments.

All of the datasets are displayed in Table 2. The twelve datasets selected were downloaded from UCI. In Table 2, $|U|$, $|C|$ and $|D|$ represent the number of objects, conditional attributes, and the categories of the decision attribute, respectively. In order to generate an incomplete information system, we deleted 12% attribute values from the raw datasets, and the missing values were randomly and uniformly distributed. The missing value of each attribute was removed with equal probability, which eliminated the impact of the later reduction set on the classification accuracy due to different attribute selection. During data preprocessing prior to the experiment, we kept the discrete data unchanged and discretized the continuous value data. The two classifiers, RF (random forest) and SMO of the Weka v3.8 software platform, were used to demonstrate the classification effect of the reduction

set. In subsequent experiments, the reduced sets of each dataset were analyzed for accuracy using the cross method. All the objects in the reduced set were randomly divided into 10 equal parts, one of which was used as the test set, and the remaining 9 were used as the training set. During the classification and analysis process, the default settings of the Weka tool were used for all parameters. In this way, each reduction set was repeated 10 times for the cross experiment. Finally, the size of the reduction set, running time, and classification accuracy obtained by the 10 experiments were averaged. We executed all experiments on a PC with Windows 10, Intel(R) Core(TM) i7-10710U CPU @ 1.10 GHz, 1.61 GHz and 8 GB memory. Algorithms were coded in python and the software that was used is PyCharm Community Edition 2020.2.3 x64.

Table 2. Description of twelve datasets.

ID	Datasets	Abbreviation	|U|	|C|	|D|
1	Promoters	Prom	106	57	2
2	Heart-statlog	Hear	270	13	2
3	Hepatitis	Hepa	155	19	2
4	HandWritten	Hand	5620	64	10
5	Chess kr-kp	Ches	3196	36	2
6	Splice	Spli	3190	61	3
7	Letters	Lett	20,000	17	26
8	Vote	Vote	435	16	2
9	Mushroom	Mush	8124	22	2
10	Qsar	Qsar	1055	42	2
11	Shuttle	Shut	43,500	9	7
12	Satimage	Sati	6435	36	6

5.1. Reduction Set Size and Running Time Analysis

This section mainly verifies the feasibility of IKAR for dataset reduction from the perspective of reduction size and calculation time. Here, we have selected three other representative incomplete system attribute reduction approaches. For the convenience of the following description, these three methods are referred to as NGLE [44], IEAR [45], and PARA [46], respectively. NGLE is the neighborhood multi-granulation rough sets-based attribute reduction using Lebesque and entropy, and the method considers both algebraic view and information view. IEAR is the information entropy attribute reduction for incomplete set-valued data using the similarity degree function and proposed λ-information entropy. PARA is the positive region attribute reduction using indiscernibility and discernibility relation.

The attribute reduction size of four algorithms is shown in Table 3. Table 4 shows the time consumed by the reduction process of the four algorithms in seconds. Ave represents the average value, Best in Table 3 represents the number of minimum reduction sets obtained, and Best in Table 4 stands for the number of times which the running time was the shortest. From Table 3, it can be seen that the average reduction set size of the IKAR algorithm is 11.833, and the average reduction set size obtained by the NGLE, IEAR, and PARA algorithms are 11.917, 11.00, and 12.083, respectively. IKAR obtained the minimum reduction set on the Ches, Spli, Mush, and Shut datasets, and the reduction effect was slightly better than the NGLE and PARA, but not as ideal as the IEAR algorithm. In the 12 datasets in Table 2, the IEAR algorithm obtains the minimum reduction set in 11 datasets.

Table 3. The attribute reduction size with the four methods on the twelve UCI datasets.

Datasets	IKAR	NGLE	IEAR	PARA
Prom	5	4	5	5
Hear	10	10	9	10
Hepa	9	10	7	10
Hand	12	11	10	11
Ches	29	30	29	30
Spli	9	10	9	9
Lett	9	9	8	9
Vote	10	9	8	10
Mush	4	5	5	5
Qsar	31	30	29	31
Shut	4	5	4	5
Sati	10	10	9	11
Ave	11.833	11.917	11.00	12.083
Best	4	1	11	1

Table 4. The run time with the four methods on the twelve UCI datasets.

Datasets	IKAR	NGLE	IEAR	PARA
Prom	0.98	2.042	3.231	1.553
Hear	0.03	0.16	2.09	0.33
Hepa	0.06	0.11	2.13	0.25
Hand	92.93	81.61	3222.89	91.28
Ches	6.35	140.98	3075.25	590.79
Spli	35.43	205.86	6557.61	309.79
Lett	3.98	6.93	117.87	8.47
Vote	0.06	0.38	6.09	1.01
Mush	4.45	411.09	4544.04	827.16
Qsar	3.84	4.97	98.76	6.61
Shut	8.98	886.08	3974.70	1133.25
Sati	25.30	125.61	2322.02	154.04
Ave	15.20	155.49	1993.89	260.38
Best	11	1	0	0

From the consumption time of Table 4, we can find that the reduction advantage obtained by the IKAR algorithm is not obvious, but that the IKAR algorithm is obviously better than the other three algorithms. When calculating the reduction set of 12 datasets, the average time required by IKAR is 15.20 s, whilst the average times consumed by the other three algorithms are (155.40, 2003.56, and 260.38) seconds, respectively. From the experimental results in Table 4, it can also be found that the IKAR algorithm only needs 25.30 s to reduce the Shut dataset, and the running time of the other three algorithms of NGLE, IEAR, and PARA is (886.08, 3974.02, and 133.25) seconds. When reducing the Mush dataset, IKAR takes 4.45 s, and the other three algorithms take (411.09, 4544.04, and 827.16) seconds. Of course, the IKAR algorithm also has shortcomings, and when the number of sample categories is higher, the algorithm is more time-consuming. For example, if there are 10 different categories of samples in the Hand dataset, the calculation time of the IKAR algorithm takes 92.93 s. At this time, the time consumption of the NGLE and PARA algorithms is (81.61 and 91.28) seconds, respectively. It takes less time than that of the IKAR algorithm.

From the above analysis, we can obtain that the IKAR algorithm can effectively reduce the dataset, it is obviously better than similar algorithms in terms of reduction speed.

5.2. Changes of the Classification Accuracy When Missing Value

It is not enough to evaluate the overall performance of a reduction algorithm only from the size and running time of the reduction set. Here, we further analyze the performance of the above four algorithms from the perspective of the classification accuracy of the reduction set. In order to find out the influence of missing values on the IKAR algorithm, we selected four datasets in Table 2, including Heas, Prom, Hepa, and Stai. During the experiment, we divided the missing values into 5 different levels, 2%, 4%, 6%, 8%, and 10%. First, 2% of the data objects in the original dataset were randomly selected, and random deletion is conducted in these data objects of attribute values for different attributes to generate an incomplete dataset. In order to reduce the bias of attribute selection that may be caused by random deletion, we used the 2% data just obtained as a basis, then selected attributes with the same probability, and then deleted the 2% data, which generated an incomplete dataset with a missing value of 4% and so on, generating a dataset with a missing value of 6%, 8%, and 10%, respectively. After the incomplete dataset was ready, we used IKAR, NGLE, IEAR, and PARA algorithms for reduction and used SMO and RF classifiers to analyze the classification accuracy. The specific results are shown in Figures 3 and 4. The horizontal axis in Figures 3 and 4 represent the proportion of deletion data objects in the dataset, and the vertical axis represents the classification accuracy.

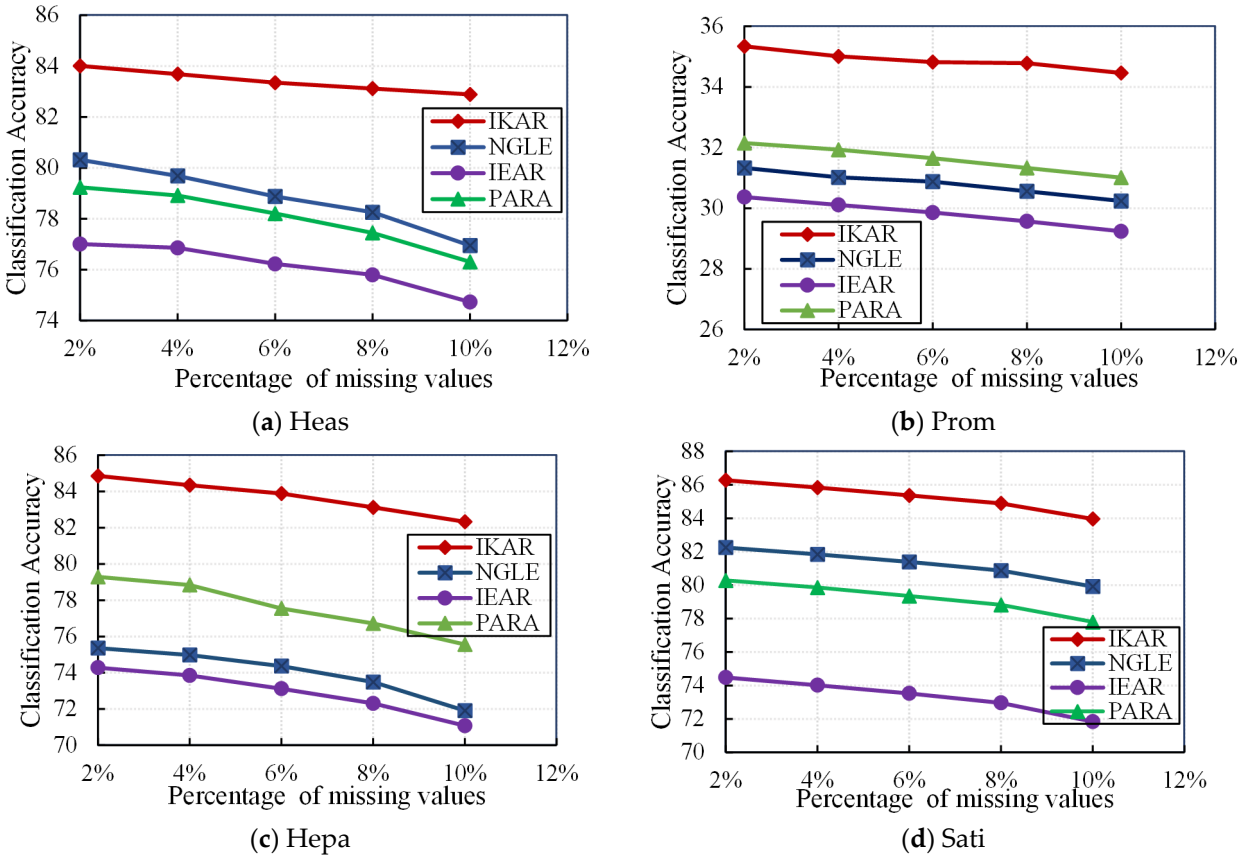

Figure 3. Variation in classification accuracy for different percentages of missing values with classifiers SMO.

Figures 3 and 4 show the change trend diagrams obtained by using the SMO and RF classifiers to analyze the accuracy of the reduced dataset. From the results of Figures 3 and 4, it can be seen that when increasing the missing data, the classification accuracy of the above four algorithms has a downward trend. For example, under the SMO classifier, when the proportion of missing values in the dataset Heas changes from 2% to 4%, the accuracy changes in IKAR and other NGLE, IEAR and PARA algorithms are (84.01→83.69),

(80.32→79.69), (77.01→76.86), and (79.24→78.92). Under the RF classifier, when the proportion of missing values in the dataset Heas changes from 2% to 4%, in the IKAR and other NGLE, IEAR, and PARA algorithms, the accuracy changes are (81.43→80.94), (77.85→76.28), (74.79→73.33), and (76.24→74.86), respectively. The main reason for this phenomenon is that as the proportion of missing data increases, more and more data objects cannot be distinguished, resulting in a decrease in classification accuracy. From the trend diagrams in Figures 3 and 4, it can be seen that the IKAR algorithm changes smoothly. The other three algorithms have a greater impact on the classification accuracy of the reduced set as the proportion of missing data increases. For example, under the RF classifier, when the missing proportion of the Hepa dataset changes from 2% to 10%, the accuracy of IKAR changes to 2.09, while the accuracy changes of the NGLE, IEAR, and PARA algorithms are 3.2, 3.19, and 4.49, respectively. Under the SMO classifier, on the classification accuracy of the Heas dataset, the IKAR algorithm changes value is 1.12, and for the other three algorithms is 3.37, 2.28, and 2.93, respectively.

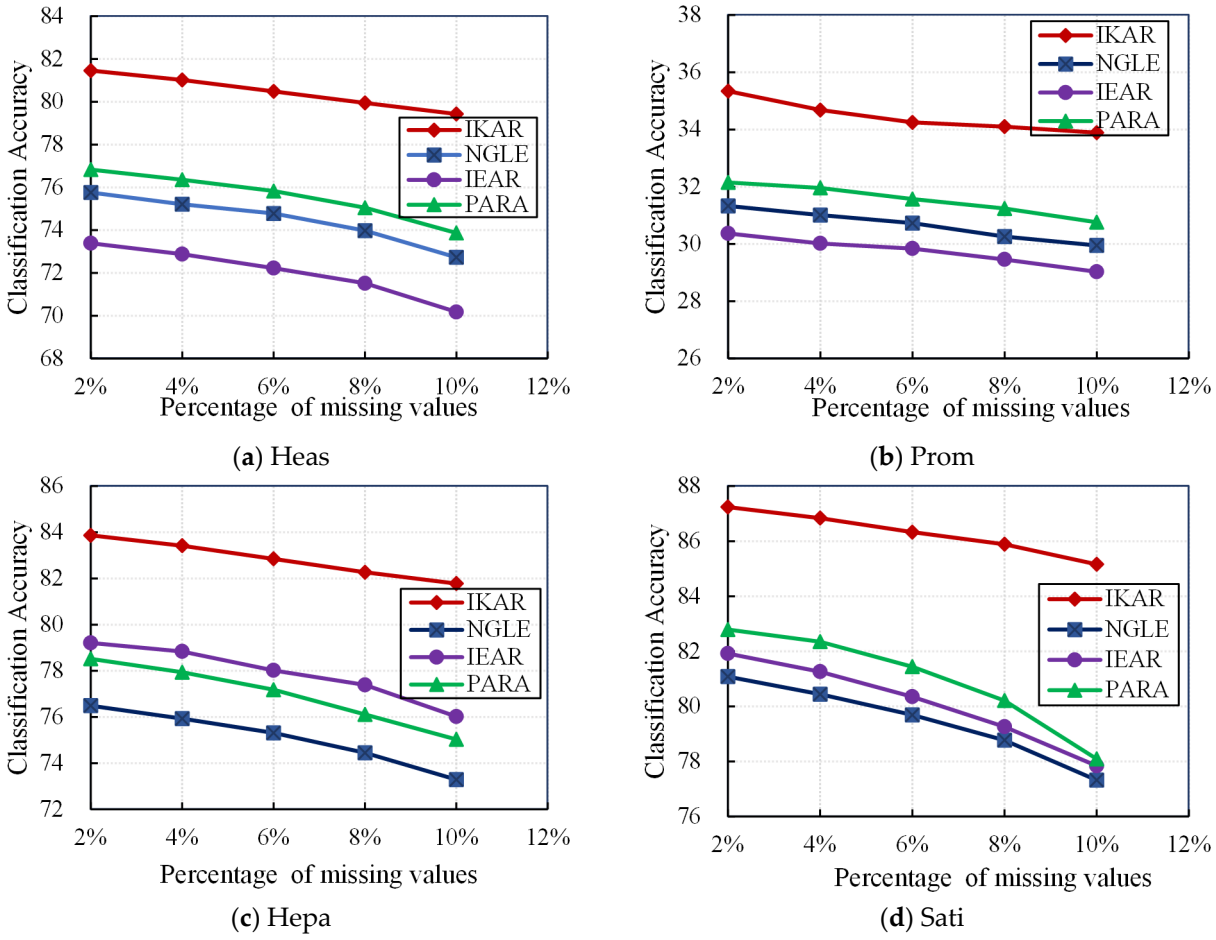

Figure 4. Variation in classification accuracy for different percentages of missing values with classifiers RF.

5.3. Classification Accuracy Analysis

The previous experiments have compared the changes in the classification accuracy when missing value exist. With the change in the proportion of missing values, the accuracy of the IKAR algorithm can not only change smoothly but can also obtain a better classification effect on multiple datasets. We used the SMO and RF classifiers to analyze the accuracy of the previous reduction set, and the detailed content is shown in Figures 5 and 6, respectively. In Figures 5 and 6, missing values represent the incomplete raw dataset and Ave denotes the average accuracy for the different datasets. Under the classifier SMO,

the IKAR algorithm obtained an average accuracy of 85.31%, and the average accuracy obtained by the other three algorithms NGLE, IEAR, PARA, and raw datasets were only 78.72, 78.78, 78.81, 84.49(%), respectively. Among the 12 datasets in Table 2, the IKAR algorithm was the highest 11 times. The classification accuracy of IKAR is higher than that of the raw set.

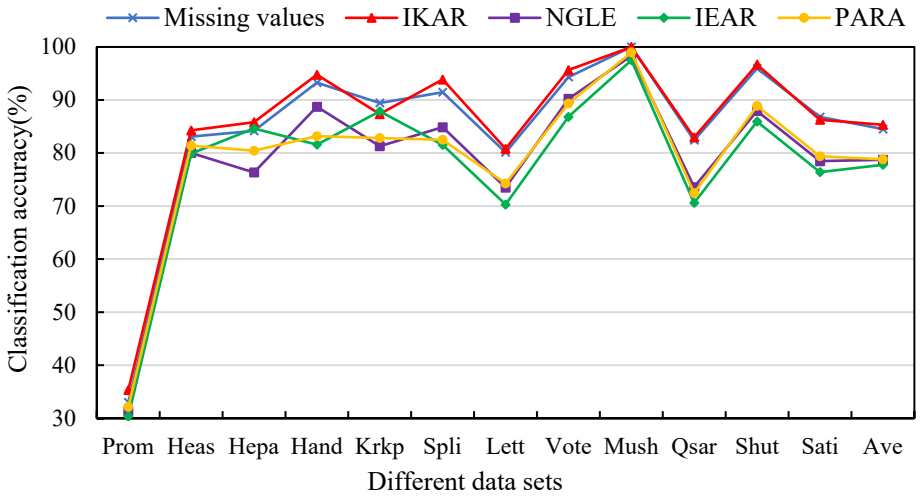

Figure 5. Trend in classification accuracy of different algorithms under the SMO classifier.

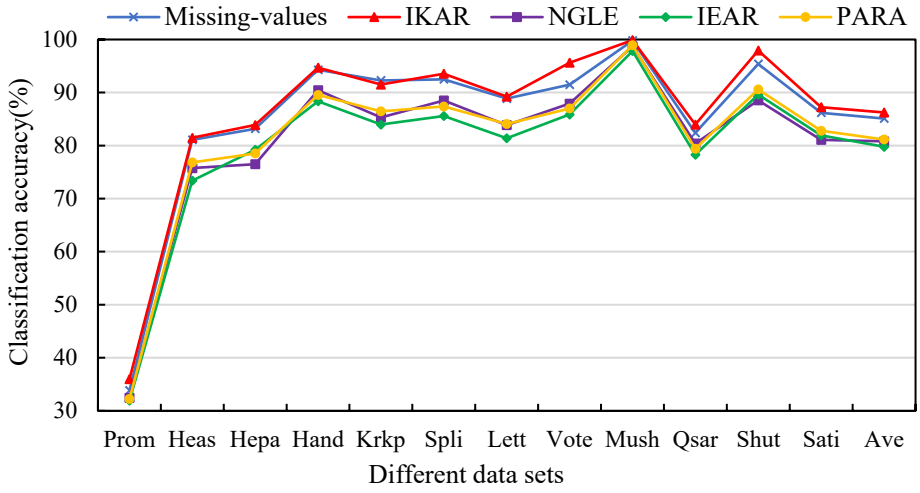

Figure 6. Trend in classification accuracy of different algorithms under the RF classifier.

For example, the classification accuracy of IKAR is (94.76, 99.68, 93.86)% in the datasets of Hand, Shut, and Spli. The classification accuracy of IEAR is (81.59, 85.93, 81.53)% and the average accuracy is 10 percentage points lower than IKAR. Similarly, on the RF classifier, the average classification accuracy of the IKAR (86.24%) is higher than the raw dataset (85.09%) and those obtained by the other three algorithms obtained (80.79, 79.77, 81.14)%, respectively.

From the results shown in Figures 5 and 6, we can obtain three conclusions as follows: (1) The classification accuracies obtained by the above four attribute algorithms are closer to those of the raw dataset, which indicates that all four algorithms are able to effectively reduce the incomplete dataset. (2) The algorithm has better overall performance. The main reason is that the IKAR algorithm considers both the correlation between attributes and the influence of attributes on the classification two factors, so it obtains a more ideal classification accuracy. (3) The classification accuracy of the IKAR algorithm is higher than the original dataset, and the main reason for this is that the IKAR algorithm effectively eliminates redundant attributes and reduces the influence of noisy data on classification.

5.4. Lgorithm Stability Analysis

To indicate the statistical significance of classification results, the non-parametric Friedman test and Bonferroni–Dunn test methods were used to the analyze classification accuracy of each classifier with different methods in Section 5.2, where the Friedman test is a statistical test that uses the ranking of each method on each dataset. The Friedman statistic is described as follows:

$$\chi_F^2 = \frac{12N}{t(t+1)} \sum_{i=1}^{s} R_i^2 - 3t(t+1) \tag{26}$$

$$F_F = \frac{(N-1)\chi_F^2}{N(t-1) - \chi_F^2} \tag{27}$$

where N is the number of datasets, and t is the categories of algorithms. R_i represents the average rank of the classification effect ranking of the i-th algorithm on all datasets, and the statistics F_F obey the Fisher distribution with $t-1$ and $(t-1)(N-1)$ degrees of freedom. If the value of F_F is bigger than $F_\alpha(t-1, N-1)$, then the original hypothesis does not hold.

The F_F test value can judge where these algorithms are different, but it cannot indicate the superiority of the algorithm. In order to explore which algorithm is better, we used the Bonferroni–Dunn test to calculate the critical value range of the average sequence value difference, which is defined as follows:

$$CD_\alpha = q_\alpha \sqrt{\frac{t(t+1)}{6N}}. \tag{28}$$

If the difference between the average ranking values of the two algorithms exceeds the critical region CD_α, the hypothesis that 'the performance of the two algorithms is the same' will be rejected with the corresponding confidence. Otherwise, the two algorithms perform differently, and the algorithm with the higher average rank is statistically better than the algorithm with the lower average rank. Generally, we set $\alpha = 0.05$.

In order to compare the stability of the IKAR algorithm in this paper, we chose three other similar algorithms, NGLE, LEAR, and PARA, to reduce the datasets in Table 2. Then, the previous reduction results were analyzed by the classifiers SMO and RF using the Weka tool. The classification accuracy was detected by the Friedman test and Bonferroni–Dunn test. When t = 4 and N = 12, then χ_F^2 = 23.4 and F_F = 20.429. Under the classifier SMO, the classification accuracy of the four algorithm reduction sets is sorted, and the number 1 represents the most ideal. The details of the sorting results are shown in Table 5. The average ranking of IKAR, NGLE, IEAR, and PARA algorithms are 1.08, 2.75, 3.58, and 2.58 in turn.

Table 5. Ranking and standard of classification accuracy under SMO classifier.

Datasets	IKAR		NGLE		IEAR		PARA	
	STD	Rank	STD	Rank	STD	Rank	STD	Rank
Prom	±0.031	1	±1.362	2	±2.019	4	±1.294	3
Heas	±0.015	1	±1.455	3	±2.318	4	±1.521	2
Hepa	±0.038	1	±2.483	4	±1.985	4	±1.984	3
Hand	±0.125	1	±2.784	2	±2.537	4	±2.183	3
Krkp	±0.247	2	±1.561	4	±1.783	1	±3.420	3
Spli	±0.035	1	±2.489	2	±3.015	4	±3.019	3
Lett	±0.512	1	±3.016	3	±1.985	4	±2.184	2
Vote	±0.187	1	±1.392	2	±2.417	4	±2.478	3
Mush	±0.261	1	±2.765	3	±2.581	4	±2.104	2
Qsar	±0.382	1	±2.054	3	±2.932	4	±3.013	2

Table 5. *Cont.*

Datasets	IKAR		NGLE		IEAR		PARA	
	STD	Rank	STD	Rank	STD	Rank	STD	Rank
Shut	±0.049	1	±1.937	3	±1.995	4	±1.329	2
Sati	±0.536	1	±2.349	3	±3.246	4	±2.586	3
AveRank		1.08		2.75		3.58		2.58

Since the critical value of $F_{0.05}(3, 33)$ is 2.892 and $F_F \succ 2.892$, we can reject the original hypothesis at $\alpha = 0.05$ under the Friedman test. So, there are statistical differences in classification accuracy among the above four algorithms. Next, $q_{0.05} = 2.569$ and $CD_{0.05} = 1.354$, and the results of Bonferroni–Dunn test for these four algorithms at $\alpha = 0.05$ is shown in Figure 7. The accuracy of the algorithm on the left side of the coordinate axis is relatively high in Figure 7. From Figure 7, we know that the average accuracy ranking of IKAR is s1 = 1.08. Among the other three algorithms, the better-ranking algorithm, PARA, has an average ranking value of s2 = 2.58. Since $|s_1 - s_2| = 1.5$ and $|s_1 - s_2| \succ CD_{0.05}$, we can find that the IKAR algorithm is significantly superior to the PARA. In the same way, the IKAR is better on accuracy than NGLE and LEAR. However, there is no obvious difference in the ranking of the NGLE, LEAR, and PARA algorithms.

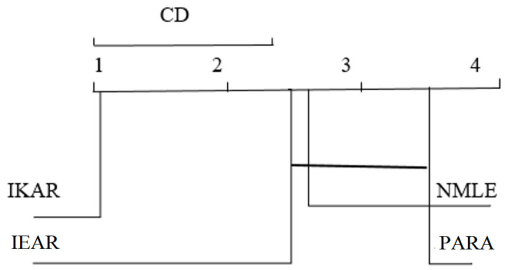

Figure 7. Bonferroni–Dunn test with SMO.

For the same reason, under the RF classifier, the average ranking values of the classification accuracy of the above four algorithms are 1.00, 2.83, 3.67, and 2.50 as shown in Table 6. Since $\chi_F^2 = 26.8$ and $F_F = 32.043$, then $F_F \succ F_{0.05}$. We can see that the classification accuracy of these four algorithms is significantly different in the statistical sense under the RF classifier. From Figure 8, we can ascertain that the IKAR algorithm is significantly different from the NGLE classification, while the ranking of the NGLE, LEAR, and PARA algorithms have no obvious differences among each other.

Table 6. Ranking and standard of classification accuracy under RF classifier.

Datasets	IKAR		NGLE		IEAR		PARA	
	STD	Rank	STD	Rank	STD	Rank	STD	Rank
Prom	±0.181	1	±1.673	2	±2.829	4	±2.639	3
Heas	±0.273	1	±2.442	3	±2.719	4	±2.521	2
Hepa	±0.357	1	±2.873	4	±2.683	2	±3.038	3
Hand	±0.362	1	±2.361	2	±3.731	4	±2.359	3
Krkp	±0.652	1	±2.048	3	±1.984	3	±2.763	2
Spli	±0.248	1	±1.994	2	±3.565	4	±3.558	3
Lett	±0.583	1	±3.246	3	±2.783	4	±3.160	2
Vote	±0.652	1	±1.817	2	±3.015	4	±2.629	3
Mush	±0.393	1	±3.092	3	±2.928	4	±1.972	2
Qsar	±0.488	1	±2.636	2	±1.997	4	±3.215	3
Shut	±0.391	1	±2.743	4	±2.963	3	±2.388	2
Sati	±0.475	1	±1.982	4	±2.837	3	±3.086	3
AveRank		1		2.83		3.67		2.5

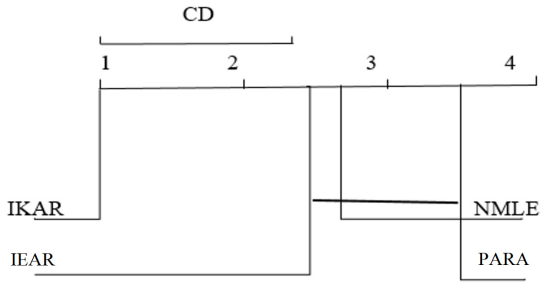

Figure 8. Bonferroni–Dunn test with RF.

In Tables 5 and 6, the STD represents the standard deviation of the classification accuracy of different algorithms' reduction datasets. From Tables 5 and 6, we know that the STD of the IKAR algorithm is smallest among the four algorithms. The STD of the classification accuracy of algorithm IKAR is less than 1 for both SMO and RF classifiers, while the STD values of the other algorithms are bigger than 1, and some are even greater than 3. For example, the STD of IKAR in the Heas dataset is ± 0.015 under the SMO, but the STD values of NGLE, LEAR, and PARA are ± 1.455, ± 2.318, and ± 1.521, respectively. On RF classifiers, the IKAR algorithm has the same stability in classification accuracy. In the dataset Spli, the STD of IKAR's classification accuracy is ± 0.248; meanwhile, the STD values of the other three algorithms are ± 1.994, ± 3.565 and ± 3.558, respectively.

Therefore, all the test results demonstrate that there is no consistent evidence to denote statistical differences between any two of the four approaches under the SMO and RF classifier. In general, the IKAR model is better than the other models in stability.

6. Conclusions

In the face of incomplete systems, most of the traditional attribute reduction models cannot obtain effective reduction rules and affect the classification accuracy of the reduced set. Therefore, we propose a new attribute reduction method, IKAR, based on the clustering background for incomplete system. IKAR uses the tolerance principle to calculate the information of knowledge granularity and to measure the similarity of data objects. When selecting attributes, the similarity of intra-cluster objects should be as large as possible, and the similarity of inter-cluster objects should be as small as possible. The innovations of this paper are manifested in the following four aspects: (1) Use of the tolerance principle to quickly calculate knowledge granularity; (2) Use of knowledge granularity to calculate the similarity of inter-cluster and intra-cluster objects; (3) Use of the idea of clustering to calculate the importance of attributes; (4) In addition to conventional time, space, and precision analysis, it also analyzes the stability of datasets with different percent missing value. All the experiments show that the IKAR algorithm is not only superior in reducing time compared to the other three algorithms, but it also has an excellent performance in terms of accuracy and stability. Of course, the IKAR algorithm also has some shortcomings. For example, it is unsuitable for datasets with multiple decision values, and complex data types and the dynamical changing of the datasets are not considered.

In our future research endeavors, we intend to work in the following four aspects. (1) We focus on attribute reduction of incomplete systems with mixed data types, especially on how to deal with missing data. (2) In addition, we will integrate the incremental learning method into the knowledge granularity reduction model in the background of clustering. (3) To address even big datasets more efficiently, applying GPU and MapReduce technologies to design parallel attribute reduction models or acceleration strategy is a very popular research topic.

Author Contributions: Conceptualization, B.L.; Supervision, R.H.; investigation, E.J.; Writing—review and editing, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of China (61836016), the Key Subject of Chaohu University (kj22zdxk01), the Quality Improvement Project of Chaohu University on Discipline Construction (kj21gczx03), and the Provincial Natural Science Research Program of Higher Education Institutions of Anhui province (KJ2021A1030).

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions, which have helped immensely in improving the quality of the paper.

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

Abbreviations

RST	Rough set theory
DS	Decision system
EGK	Equal knowledge granularity
GK	Knowledge granularity
$SInter_{Di}(P)$	The similarity of objects in the clustering of Di under attribute set P
ASIntra	The average intra-cluster similarity
SInter	The inter-cluster similarity
AInter	The average inter-cluster similarity
SIM	The importance of attribute
Sig	The inner significance
IKAR	The knowledge granularity attribute reduction for incomplete systems
NGLE	Neighborhood multi-granulation attribute reduction using Lebesque and Entropy
PARA	The positive attribute reduction algorithm
IEAR	The information entropy attribute reduction

References

1. Pawlak, Z. Rough sets. *Int. J. Comput. Inform. Sci.* **1982**, *11*, 341–356. [CrossRef]
2. Wang, R.; Wang, X.; Kwong, S.; Xu, C. Incorporating diversity and informative-ness in multiple-instance active learning. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1460–1475. [CrossRef]
3. Wang, X.; Tsang, E.C.C.; Zhao, S.; Chen, D.; Yeung, D.S. Learning fuzzy rules from fuzzy samples based on rough set technique. *Inf. Sci.* **2017**, *177*, 4493–4514. [CrossRef]
4. Wang, X.; Xing, H.; Li, Y.; Hua, Q.; Dong, C.R.; Pedrycz, W. A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning. *IEEE. Trans. Fuzzy. Syst.* **2015**, *23*, 638–1654. [CrossRef]
5. Liu, X.; Qian, Y.; Liang, J. A rule-extraction framework under multi-granulation rough sets. *Int. J. Mach. Learn. Cybern.* **2014**, *5*, 319–326. [CrossRef]
6. Zhang, X.; Mei, C.; Chen, D.; Li, J. Multi-confidence rule acquisition and confidence-preserved attribute reduction in interval valued decision systems. *Int. J. Approx. Reason.* **2014**, *55*, 1787–1804. [CrossRef]
7. Hu, Q.; Yu, D.; Xie, Z. Neighborhood classifiers. *Expert Syst. Appl.* **2008**, *34*, 866–876. [CrossRef]
8. Cheruku, R.; Edla, D.R.; Kuppili, V.; Dharavath, R. RST-Bat-Miner: A fuzzy rule miner integrating rough set feature selection and bat optimization for detection of diabetes disease. *Appl. Soft. Comput.* **2018**, *67*, 764–780. [CrossRef]
9. Hamouda, S.K.M.; Wahed, M.E.; Alez, R.H.A.; Riad, K. Robust breast cancer prediction system based on rough set theory at National Cancer Institute of Egypt. *Comput. Methods Programs Biomed.* **2018**, *153*, 259–268. [CrossRef]
10. Jothi, G.; Inbarani, H.H. Hybrid tolerance rough set-firefly based supervised feature selection for MRI brain tumor image classification. *Appl. Soft. Comput.* **2016**, *46*, 639–651.
11. Hao, C.; Li, J.; Fan, M.; Liu, W.; Tsang, E.C. Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions. *Inf. Sci.* **2017**, *415*, 213–232. [CrossRef]
12. Liang, D.; Xu, Z.; Liu, D. Three-way decisions based on decision-theoretic rough sets with dual hesitant fuzzy information. *Inf. Sci.* **2017**, *396*, 127–143. [CrossRef]
13. Qu, J.; Bai, X.; Gu, J.; Taghizadeh-Hesary, F.; Lin, J. Assessment of Rough set Theory in Relation to Risks Regarding Hydraulic Engineering Investment Decisions. *Mathematics* **2020**, *8*, 1308. [CrossRef]
14. Lei, L. Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. *Appl. Soft Comput.* **2018**, *62*, 923–932. [CrossRef]

15. Singh, A.K.; Baranwal, N.; Nandi, G.C. A rough set based reasoning approach for criminal identification. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 413–431. [CrossRef]
16. Lin, Y.; Li, J.; Lin, P.; Lin, G.; Chen, J. Feature selection via neighborhood multi-granulation fusion. *Knowl. Based Syst.* **2014**, *67*, 162–168. [CrossRef]
17. Sun, L.; Xu, J. A granular computing approach to gene selection. *Bio-Med. Mater. Eng.* **2014**, *24*, 1307–1314. [CrossRef]
18. Fujita, H.; Gaeta, A.; Loia, V.; Orciuoli, F. Resilience Analysis of Critical Infrastructures: A cognitive approach based on Granular Computing. *IEEE Trans. Cybern.* **2018**, *49*, 1835–1848. [CrossRef]
19. Qian, H.Y. Granulation Mechanism and Data Modeling of Complex Data. Ph.D. Thesis, Shanxi University, Taiyuan, China, 2011.
20. Dai, J.; Wang, W.; Zhang, C.; Qu, S. Semi-supervised attribute reduction via attribute indiscernibility. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 1445–1464. [CrossRef]
21. Gao, C.; Wang, Z.C.; Zhou, J. Three-way approximate reduct based on information-theoretic measure. *Int. J. Approx. Reason.* **2022**, *142*, 324–337. [CrossRef]
22. Yang, L.; Qin, K.; Sang, B.; Fu, C. A novel incremental attribute reduction by using quantitative dominance-based neighborhood self-information. *Know. Based Syst.* **2023**, *261*, 110200. [CrossRef]
23. Wang, Z.H.; Zhang, X.P. The granulation attribute reduction of multi-label data. *Appl. Intell.* **2023**, *53*, 19266–19284. [CrossRef]
24. Yu, B.; Hu, Y.; Kang, Y.; Cai, M. A novel variable precision rough set attribute reduction algorithm based on local attribute significance. *Int. J. Approx. Reason.* **2023**, *157*, 88–104. [CrossRef]
25. Zhang, S.C.; Cheng, D.B.; Hu, R.Y.; Deng, Z. Supervised feature selection algorithm via discriminative ridge regression. *World Wide Web* **2018**, *21*, 1545–1562. [CrossRef]
26. Liu, T.; Hu, R.Y.; Zhu, Y.X. Completed sample correlations and feature dependency-based unsupervised feature selection. *Multim. Tools Appl.* **2023**, *82*, 15305–15326. [CrossRef]
27. Devi, P.; Kizielewicz, B.; Guleria, A.; Shekhovtsov, A.; Gandotra, N.; Saini, N.; Sałabun, W. Dimensionality reduction technique under picture fuzzy environment and its application in decision making. *Int. J. Knowl. Based Intell. Eng. Syst.* **2023**, *27*, 87–104. [CrossRef]
28. Wen, H.T.; Zhao, S.X.; Liang, M.S. Unsupervised attribute reduction algorithm for mixed data based on fuzzy optimal approximation set. *Mathematics* **2023**, *11*, 3452. [CrossRef]
29. Li, Z.; Liao, S.; Qu, L.; Song, Y. Attribute selection approaches for incomplete interval-value data. *J. Intell. Fuzzy Syst.* **2021**, *40*, 8775–8792. [CrossRef]
30. Liu, X.F.; Dai, J.H.; Chen, J.L.; Zhang, C.C. A fuzzy α-similarity relation-based attribute reduction approach in incomplete interval-valued information systems. *Appl. Soft Comput.* **2021**, *109*, 107593. [CrossRef]
31. Dai, J.H.; Wang, Z.Y.; Huang, W.Y. Interval- valued fuzzy discernibility pair approach for attribute reduction in incomplete interval-valued information systems. *Inf. Sci.* **2023**, *642*, 119215. [CrossRef]
32. Song, Y.; Luo, D.; Xie, N.; Li, Z. Uncertainty measurement for incomplete set-valued data with application to attribute reduction. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 3031–3069. [CrossRef]
33. Zhou, Y.; Bao, Y.L. A Novel Attribute Reduction Algorithm for Incomplete Information Systems Based on a Binary Similarity Matrix. *Symmetry* **2023**, *15*, 674. [CrossRef]
34. Zhang, C.L.; Li, J.J.; Lin, Y.D. Knowledge reduction of pessimistic multigranulation rough sets in incomplete information systems. *Soft Comput.* **2021**, *25*, 12825–12838. [CrossRef]
35. He, J.; Qu, L.; Wang, Z.; Chen, Y.; Luo, D.; Wen, C.F. Attribute reduction in an incomplete categorical decision information system based on fuzzy rough sets. *Artif. Intell. Rev.* **2022**, *55*, 5313–5348. [CrossRef]
36. Srirekha, B.; Sathish, S.; Devi, R.M. Attribute reduction on SE-ISI Concept Lattice for an Incomplete Context using object ranking. *Mathematics* **2023**, *11*, 1585. [CrossRef]
37. Cornelis, C.; Jensen, R.; Martin, G.H.; Slezak, D. Attribute selection with fuzzy decision reducts. *Inf. Sci.* **2010**, *180*, 209–224. [CrossRef]
38. Liu, G.L.; Feng, Y.B.; Yang, J.T. A common attribute reduction form for information systems. *Know. Based Syst.* **2020**, *193*, 105466. [CrossRef]
39. Nguyen, N.T.; Sartra, W. A novel feature selection Method for High-Dimensional Mixed decision tables. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 3024–3037.
40. Zhang, M.; Chen, D.G.; Yang, Y.Y. A new algorithm of attribute reduction based on fuzzy clustering. In Proceedings of the International Conference on Machine Learning and Cybernetics, Tianjin, China, 14–17 July 2013; Volume 7, pp. 14–17.
41. Jia, H.J.; Ding, S.F.; Ma, H.; Xing, W.Q. Spectral Clustering with Neighborhood Attribute Reduction Based on Information. Entropy. *J. Comput.* **2014**, *9*, 1316–1324. (In Chinese) [CrossRef]
42. Zhao, R.N.; Gu, L.Z.; Zhu, X.N. Combining Fuzzy C-Means Clustering with Fuzzy Rough Feature Selection. *Appl. Sci.* **2019**, *9*, 679. [CrossRef]
43. Jia, X.Y.; Rao, Y.Y.; Shang, L.; Li, T.G. Similarity-based attribute reduction in rough set theory: A clustering perspective. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 1047–1060. [CrossRef]
44. Zhang, Y.; Chen, Z. Information Entropy-Based Attribute Reduction for Incomplete Set-Valued Data. *IEEE Access* **2021**, *10*, 8864–8882. [CrossRef]

45. Shu, W.H.; Qian, W.B. A fast approach to attribute reduction from perspective of attribute measures in incomplete decision systems. *Know. Based Syst.* **2014**, *72*, 60–71. [CrossRef]
46. Sun, L.; Wang, L.; Ding, W.; Qian, Y.; Xu, J. Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems. *Know. Based Syst.* **2020**, *192*, 105373. [CrossRef]