

Article

Attribute Reduction Based on LIFT and Random Sampling

Qing Chen ¹, Taihua Xu ^{1,2,*} and Jianjun Chen ¹¹ School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212100, China² Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province, Zhejiang Ocean University, Zhoushan 316022, China

* Correspondence: xth19890410@163.com

Abstract: As one of the key topics in the development of neighborhood rough set, attribute reduction has attracted extensive attentions because of its practicability and interpretability for dimension reduction or feature selection. Although the random sampling strategy has been introduced in attribute reduction to avoid overfitting, uncontrollable sampling may still affect the efficiency of search reduct. By utilizing inherent characteristics of each label, Multi-label learning with Label specific Features (LIFT) algorithm can improve the performance of mathematical modeling. Therefore, here, it is attempted to use LIFT algorithm to guide the sampling for reduce the uncontrollability of sampling. In this paper, an attribute reduction algorithm based on LIFT and random sampling called ARLRS is proposed, which aims to improve the efficiency of searching reduct. Firstly, LIFT algorithm is used to choose the samples from the dataset as the members of the first group, then the reduct of the first group is calculated. Secondly, random sampling strategy is used to divide the rest of samples into groups which have symmetry structure. Finally, the reducts are calculated group-by-group, which is guided by the maintenance of the reducts' classification performance. Comparing with other 5 attribute reduction strategies based on rough set theory over 17 University of California Irvine (UCI) datasets, experimental results show that: (1) ARLRS algorithm can significantly reduce the time consumption of searching reduct; (2) the reduct derived from ARLRS algorithm can provide satisfying performance in classification tasks.



Citation: Chen, Q.; Xu, T.; Chen, J. Attribute Reduction Based on LIFT and Random Sampling. *Symmetry* **2022**, *14*, 1828. <https://doi.org/10.3390/sym14091828>

Academic Editor: Mihai Postolache

Received: 15 August 2022

Accepted: 1 September 2022

Published: 3 September 2022

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



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

Keywords: attribute reduction; LIFT; neighborhood rough set; random sampling

1. Introduction

In data processing, as one of the mechanisms to obtain important attributes, attribute reduction has been widely studied. In general, the goal of attribute reduction is to eliminate unnecessary or irrelevant attributes from the raw attributes in light of the given constraint. The remaining attributes are then referred to as a reduct. Then, how to explore attribute reduction has experienced a long development.

In the early stage of exploring attribute reduction [1–4], the exhausted searching strategies such as the discernibility matrix-based strategy [5–7] grew vigorously. Unfortunately, in the age of big data, such time consumption becomes intolerable as data volume grows. This is mostly due to the fact that the majority of exhausted searches are NP-hard issues [8]. For these considerations, various heuristic searching-based strategies have been designed [9]. For example, Liu et al. [10] presented the bucket in the searching process. It can minimize the time necessary to derive reduct by reducing the calculations of distance between samples. Qian et al. [11] proposed an strategy based on positive rough set approximation that may significantly reduce sample size in the searching method. Clearly, both of them are designed from the perspective of samples, which means the searching efficiency is improved by compressing the space of samples.

From this point of view, it is not difficult to confirm that the efficiency of these two strategies depend strongly on the sample distribution. For example, the iterative efficiency of positive approximation decreases when the constrain of attribute reduction is difficult

to met. Even the positive approximation will also lose its effectiveness, because the more iterations attribute reduction has, the fewer samples can be removed in each iteration. As far as the bucket is concerned, the samples may all appear in the same bucket in the worst case. In this case, the bucket would be invalid, even redundant, because it is not only necessary to calculate the distance between any two samples, but also requires additional time to obtain the bucket.

In order to reduce the strong influence of sample distribution on attribute reduction algorithm. Chen et al. [12] developed a random sampling accelerator. The first step of random sampling accelerator is to divide the samples into several groups, which can lead to rapid attribute assessment and selection for each group. The second step is to calculate reduct group-by-group based on guidance, which can reduce the search space of candidate attributes. Then, there are fewer candidate attributes to be evaluated, so the calculation process can be accelerated. However, the first group of samples constructed by Chen et al. [12] is still affected by the sample distribution. Even though the size of the first group of samples is much smaller than the entire dataset, there is often overfitting [13–15] in the first group, which will inevitably affect the reduction of subsequent groups.

In order to solve the above problem, it is needed to find the most representative samples of different labels to avoid overfitting. Multi-label learning with Label specific FeaTures (LIFT) [16] has been proved to be effective for above needs. Briefly, LIFT algorithm constructs inherent characteristics to every label by conducting clustering analysis on its negative class samples and positive class samples. Therefore, we try to optimize random sampling process by utilization of LIFT algorithm, that is, the centers obtained from each cluster are used as the members of first group.

In this paper, the random sampling strategy optimized by LIFT algorithm includes three main stages: (1) extract the samples that can best distinguish each label; (2) randomly group the rest of samples in the dataset (the dataset is a finite set of all samples); (3) the samples selected in the first step are trained, and the corresponding reduct is then obtained. For the reduction process of subsequent groups, the reduct of the previous groups is used at the current group to test the classification performance. If the given constraint is met, the reduct of this group will not be calculated. If not, the previously trained samples will be added into current group. Then, it is continued to calculate the reduct of the current group.

In the following, the main contributions of the proposed will be detailed: (1) By using LIFT, the influence of sample distribution is further reduced, which speeds up the subsequent reduction and has satisfying classification performance. (2) Through random sampling strategy, the sample volume is reduced to obtain higher time efficiency. In addition to the two advantages mentioned above, it should be noted that our strategy may also be easily integrated with other well-liked strategies. For example, the core frameworks of the bucket and the positive approximation are easily inserted into our construct for increasing the efficiency of searching reducts further.

The remainder of this paper is structured as follows: Section 2 discusses the basic concepts used in this paper. Section 3 describes the optimization of the random sampling strategy by LIFT algorithm. Section 4 reports the comparative experimental results and detailed analysis. Section 5 summarizes the contributions of this paper and arranges some further prospects.

2. Preliminaries

2.1. Basic Concept of Rough Set

Usually, a decision system can be described as $\mathcal{DS} = \langle U, AT \cup \{d\} \rangle$: the universe U is the finite set including all samples; AT is the finite set of condition attributes; d is a label. $\forall x_i \in U, \forall a \in AT, a(x_i)$ implies the attribute value of sample x_i over the attribute a , $d(x_i)$ indicates the label of sample x_i . Immediately, if $A \subseteq AT$, then an indiscernibility relation can be given by $IND_A = \{(x_i, x_j) \in U \times U : a(x_i) = a(x_j), \forall a \in A\}$ which is also an equivalence relation having reflexivity, symmetry, transitivity. As a result, the construction of classical rough set will be given in the Definition 1.

Definition 1. Given a decision system \mathcal{DS} , $\forall A \subseteq AT, \forall X \subseteq U$, the lower and upper approximations of X related to A are defined as:

$$\underline{R}X = \{x_i \in U : \text{IND}_A(x_i) \subseteq X\}; \quad (1)$$

$$\overline{R}X = \{x_i \in U : \text{IND}_A(x_i) \cap X \neq \emptyset\}; \quad (2)$$

$\text{IND}_A(x_i) = \{x_j \in U : (x_i, x_j) \in \text{IND}_A\}$ is the set of all samples in U which are equivalent to x in terms of A .

According to Definition 1, the equivalence class $\text{IND}_A(x)$ is the key to construct upper and lower approximations. In general, the process of obtaining $\text{IND}_A(x)$ is known as information granulation in the field of Granular Computing. Although indiscernibility relation is used to perform information granulation in Definition 1, other strategies of dealing with complex data have also been developed, e.g., information granulation based on fuzzy [17–20], information granulation based on neighborhood [21,22], information granulation based on clustering, information granulation based on strongly connected components [23] and so on [24].

It should be emphasized that the information granulation based on neighborhood is especially accepted in many researching fields. Such a mechanism is not only suitable to mixed data, but also equipped with the natural structure of multi-granularity. The detailed form is given as follows.

Definition 2. Given a decision system \mathcal{DS} , $\forall A \subseteq AT, \forall x_i \in U$, the neighborhood $\delta_A(x_i)$ is defined as:

$$\delta_A(x_i) = \{x_j : x_j \in U, \text{dis}_A(x_i, x_j) \leq \delta\}, \quad (3)$$

in which $\text{dis}_A(x_i, x_j)$ indicates the distance between sample x_i and sample x_j in the terms of A , δ is a given radius.

$\delta_A(x_i)$ indicates the neighborhood based information granule centered with sample x_i in the terms of A . As a result, Definition 3 can be used to denote the definitions of upper and lower approximations in the neighborhood rough set.

Definition 3. Given a decision system \mathcal{DS} , $\forall A \subseteq AT, \forall X \subseteq U$, the neighborhood-based lower and upper approximations of X are defined as:

$$\underline{N}X = \{x_i : x_i \in U, \delta_A(x_i) \subseteq X\}; \quad (4)$$

$$\overline{N}X = \{x_i : x_i \in U, \delta_A(x_i) \cap X \neq \emptyset\}. \quad (5)$$

Following Definition 3, Approximation Quality (AQ) is a definition of a measure for expressing the degree of approximation in terms of the neighborhood rough set, which will show in the following:

Definition 4. Given a decision system \mathcal{DS} , $\forall A \subseteq AT$, the approximation quality in terms of neighborhood rough set is:

$$\text{AQ}(A) = \frac{|\{x_i : x_i \in U, \delta_A(x_i) \subseteq [x_i]_d\}|}{|U|}, \quad (6)$$

in which $[x_i]_d = \{x_j : x_j \in U, d(x_i) = d(x_j)\}$ indicates the label which contains x_i .

2.2. Attribute Reduction

At present, not only the amount, but also the dimensions of datasets are quickly increasing. In the realm of rough set, an interesting mechanism known as attribute reduction can efficiently reduce dimension by deleting superfluous attributes. Various kinds of

attribute reduction strategies have been described for various needs [25–31]. However, Yao et al. [22] noted that the majority of them have a similar mechanism. They provided the following general form to clearly express the concept of attribute reduction.

Definition 5. Given a decision system \mathcal{DS} , \mathcal{C}_ρ^U is a constraint related to measure ρ over the universe U , $\forall A \subseteq AT$, A is referred to as a \mathcal{C}_ρ^U -reduct if and only if the following conditions hold:

- (1) A meets the constraint \mathcal{C}_ρ^U ;
- (2) $\forall A' \subset A$, A' does not meet the constraint \mathcal{C}_ρ^U .

In Definition 5, ρ can be regarded as a function such that $\rho : P(U) \times P(AT) \rightarrow \mathbb{R}$, \mathbb{R} is the set of all real numbers, $P(U)$ and $P(AT)$ are the power sets of U and AT , respectively. For instance, $\forall x_i \in U$ and $\forall a \in AT$, if the lower approximation is employed, then by using a , $\rho(x_i, a) = 1$ or 0 means that x_i is in or out of the lower approximation. Some typical explanations will be displayed in Example 1 below.

Example 1. The following two phases can be elaborated.

- (1) The greater the degree of the constrain obtained by rho, the less the level of approximation will be described. One example of such a metric is the definition of approximation quality.
- (2) The less the degree of the constrain we derived by utilizing rho, the less the amount of approximation in the dataset. One example of such a metric is the idea of decision error rate.

Finally, $\forall U' \subseteq U$, $\rho(U', A)$ can be considered as a fusion of $\{\rho(x_i, A) : \forall x_i \in U'\}$, is denoted by

$$\rho(U', A) = \bigodot_{x_i \in U'} \rho(x_i, A). \quad (7)$$

In Equation (7), \odot is viewed as a fusion operator [32,33] with a variety of forms. In [32], the fusion operator is explained as the averaging operator. In addition, in [33], the fusion operator is explained as the summation operator. As a result, various forms to calculate $\rho(U', A)$ can yield various results. For instance, if \odot can be demonstrated at \sum , i.e., $\rho(U', A) = \sum_{x_i \in U'} \rho(x_i, A)$, then $\rho(U', A)$ can be utilized to determine the degree of approximation quality over U' .

2.3. Obtaining Reduct

Following Definition 5, the minimal attribute subset is called a reduct which meet the constraint \mathcal{C}_ρ^U . Immediately, it has become a serious challenge how to seek out such the minimum attribute subset. Algorithm 1 depicts a comprehensive procedure of forward greedy searching that can be utilized to determine one reduct.

For every iteration of determining reduct in Algorithm 1, every candidate attribute must be examined. In the worst-case, every attribute in AT is added to the reduct. As a result, the time complexity of Algorithm 1 is $\mathcal{O}(|U|^2 \cdot |AT|^2)$.

Even if the Algorithm 1 is successful in obtaining reduct, there are still problems. For instance, if the size of dataset becomes large, the consumption of time to derive reduct will be too extensive. As a result, several researchers have built many strategies from diverse perspectives. All of them have their own disadvantages and advantages. In the next subsection, three strategies built from the perspective of samples will be introduced in detail.

Algorithm 1: Forward Greedy Searching for Attribute Reduction (FGSAR)

Input: Decision system \mathcal{DS} , constraint \mathcal{C}_ρ^U .
Output: One reduct A .

- 1 Perform information granulation over U by using AT ;
- 2 Calculate the measure-value such that $\rho(U, AT) = \bigcirc_{x \in U} \rho(x, AT)$;
- 3 $A = \emptyset$;
- 4 **While** \mathcal{C}_ρ^U is not met **do**
- 5 // Determine the required attributes
- 6 $\forall a \in AT - A$, perform information granulation over U by using $A \cup \{a\}$;
- 7 Analyze a according to compute $\rho(U, A \cup \{a\})$;
- 8 Choose a an appropriate attribute $b \in AT - A_i$ by the standard;
- 9 $A = A \cup \{b\}$;
- 10 Calculate $\rho(U, A)$;
- 11 **End**
- 12 **Repeat**
- 13 // Eliminate superfluous attributes
- 14 $\forall c \in A$, perform information granulation over U by using $A - \{c\}$;
- 15 Calculate $\rho(U, A - \{c\})$;
- 16 **If** \mathcal{C}_ρ^U is met
- 17 | $A = A - \{c\}$;
- 18 **End**
- 19 **Until** A remains unchanged;
- 20 // $|A|$ represents the fundamental amount of set A
- 21 **Return** A .

2.4. Strategy Based on Sample

The key of strategy based on sample is compression of sample space [34]. These strategies can immediately reduce the number of scanned samples or comparisons between samples. Here are three classical strategies based on the sample.

- *Bucket based searching for attribute reduction.* Liu et al. [10] have considered the bucket method for fast obtaining reduct. A hash function is used in their technique, and every sample in the dataset will then be mapped into a number of separate buckets. Following the intrinsic properties of such a mapping, only samples from the same bucket must be compared rather than samples from the entire dataset. According to this viewpoint, the computing burden of information granulation may be minimized. As a result, the technique for obtaining reduct by the bucket-based strategy is identical to that of Algorithm 1, except for the device of information granulation.
- *Positive approximation for attribute reduction (PAAR).* Qian et al. [11] have presented the method based on positive approximation to calculate reduct rapidly. The essence to positive approximation is to compress gradually the sample space. The compressing process should be guided by the values associated with the constrain ρ . The following are the precise steps of the attribute reduction based on positive approximation.
 - (1) The hypothetical reduct A will be defined as \emptyset , as well as the sample compressed space U' will be defined as U .
 - (2) By using the constrain of $\rho(U', A \cup \{a\})$, analyse all hypothetical $a \in AT - A$ which in U' .
 - (3) Using the acquired constrain, choose one qualified attribute $b \in AT - A$ then merge in b to A .
 - (4) Based on A , calculate $\rho(U', A)$ and then reanalysis construction of the sample compressed space U' .
 - (5) If the specified constraint is met, produce reduct A ; otherwise, back to step(2).

- *Random sampling accelerator for attribute reduction (RSAR)*. Chen et al. [12] examined the above-mentioned strategies, they found that: (1) Regardless of the searching technique, information granulation over the entire dataset is required; (2) Information granulation over the dataset always needs to be regenerated in each iteration throughout the entire searching process. In this aspect, sample distribution may influence the effectiveness of searching. Therefore, the two above-mentioned strategies have their own restrictions since they are directly tied to sample distribution. The restriction of BBSAR is that Bucket strategy will become inefficient when sample is too centralized, which is time-consuming. The restriction of PAAR is that the sample distribution strongly affect the construction of positive approximation, and then affect the effectiveness of attribute reduction. In view of this, Chen et al. [12] developed a new random sampling strategy. The following shows the exact structure of the random sampling being used to derive reduct.
 - (1) The samples were randomly separated to n sample groups of equal size: U_1, \dots, U_n .
 - (2) Compute the reduct A_1 over U_1 ; reduct A_1 will then provide advice for computing the reduct A_2 over $U_1 \cup U_2$, and so on.
 - (3) Get the reduct A_n , use it as the ultimate reduct over the entire dataset.

RSAR can reduce the influence of sample distribution on attribute reduction to a certain extent. It is well-known that the reduct obtained from a small volume of samples often does not represent the attribute importance of the entire dataset. However, there are two facts in RSAR: (1) the volume of first sample group is frequently much less than the volume of entire dataset; (2) the reduct obtained from the first sample group is used to guide the reduction calculation of subsequent sample groups which have symmetry. These two facts may reduce the efficiency of attribute reduction. As a result, developing an effective attribute reduction strategy to solve the problem is widely valued.

3. LIFT for Attribute Reduction

3.1. Theoretical Foundations

This section provides an attribute reduction strategy based on random sampling and LIFT [16] which can select a small volume of represented samples from entire dataset. LIFT considers that samples with different labels should have their own characteristics, which should be helpful to find samples with high discernibility.

In LIFT, given an n -dimensional attribute space, $DL = \{d^1, d^2, \dots, d^q\}$ represents the finite set of q possible labels possessed by all samples. Therefore, the sample space can be expressed as $S = \{(x_i, d_i) \mid i = 1, 2, \dots, m\}$, x_i is a sample with n -dimensional attribute in the dataset, and $d_i \subseteq DL$ represents the label set of x_i . In order to extract the characteristics that can best distinguish one label from every other label, the internal relationship between samples under each label will be consider. Specifically, for a label $d^k \in DL$, samples are divided into two categories, the positive class samples set P_k and the negative class samples set N_k , as shown in following:

$$P_k = \{x_i \mid (x_i, d_i) \in S, d^k \in d_i\}; \quad (8)$$

$$N_k = \{x_i \mid (x_i, d_i) \in S, d^k \notin d_i\}. \quad (9)$$

For the label d^k , if one sample possesses this label, the sample will be classified into positive class, otherwise it will be classified into negative class. In order to explore the internal structures of positive and negative classes, k -means [35] clustering is, respectively, carried out on P_k and N_k . Since the clustering information of P_k can be regarded as equally important with N_k , the number of cluster centers is set as m_k , that is:

$$m_k = \lceil \delta \cdot \min(|P_k|, |N_k|) \rceil; \quad (10)$$

$|\cdot|$ is the cardinality of set; $\delta \in [0, 1]$ is the parameters controlling the number of clusters; the cluster centers of P_k and N_k are recorded as $\{p_1^k, p_2^k, \dots, p_{m_k}^k\}$ and $\{n_1^k, n_2^k, \dots, n_{m_k}^k\}$.

By utilizing the ability of LIFT on selecting the represented samples, we try to give a new strategy for deriving reduct:

- (1) Collect all cluster centers into U_1 .
- (2) The other samples were randomly separated to $n - 1$ sample groups of equal size: U_2, \dots, U_n .
- (3) Compute the first reduct A_1 over U_1 ; this reduct will then provide advice for computing the second reduct A_2 over $U_1 \cup U_2$; A_2 will also provide advice for computing A_3 over $U_1 \cup U_2 \cup U_3$ and so on.
- (4) Get the n -th reduct A_n , use it as the ultimate reduct over the entire dataset.

According to the previous description, it can be seen that the new strategy provides the following advantages. To begin, when the sample distribution information is unclear, LIFT is used to select the most representative samples as the first sample group to better guide the subsequent attribute reduction. Then, except the attribute reduction of the first sample group, the sample distribution will no longer affect the efficiency of attribute reduction of other sample groups which have symmetry. Obviously, these two advantages may aid in reducing the time consumption of attribute reduction.

3.2. Detailed Algorithm

In the worst case of the Algorithm 1, it contains two aspects: (1) There is always only one attribute to be added to the reduct in the each iteration; (2) The final reduct contains all attributes in the dataset. After that, the following will show the time complexity of Algorithm 2.

Remark 1. The time complexity of Algorithm 2 is $\mathcal{O}\left(\left(\frac{1}{n}|U|\right)^2 \cdot S_1^2 + \left(\frac{2}{n}|U|\right)^2 \cdot S_2^2 + \dots + \left(\frac{n}{n}|U|\right)^2 \cdot S_n^2\right)$, i.e., $\mathcal{O}\left(\frac{|U|^2}{n^2} \cdot \sum_{i=1}^n (i^2 \cdot S_i^2)\right)$. For such complexity, n refers to the amount of sample groups, S_n indicates the amount attributes ought to be included in the reduct for every iterations, thus $\sum_{i=1}^n S_i = S_1 + S_2 + \dots + S_n = |AT|$. As a result, it is simple to arrive at the following conclusions:

$$\begin{aligned} & \mathcal{O}\left(\left(\frac{1}{n}|U|\right)^2 \cdot S_1^2 + \left(\frac{2}{n}|U|\right)^2 \cdot S_2^2 + \dots + \left(\frac{n}{n}|U|\right)^2 \cdot S_n^2\right) \\ &= \mathcal{O}\left(\frac{|U|^2}{n^2} \cdot \sum_{i=1}^n (i^2 \cdot S_i^2)\right) \\ &\leq \mathcal{O}\left(\frac{|U|^2}{n^2} \cdot \sum_{i=1}^n (n^2 \cdot S_i^2)\right) \\ &= \mathcal{O}\left(|U|^2 \cdot \left(\sum_{i=1}^n S_i^2\right)\right) \\ &\leq \mathcal{O}\left(|U|^2 \cdot \left(\sum_{i=1}^n S_i\right)^2\right), \end{aligned}$$

$$\text{i.e., } \mathcal{O}\left(\frac{|U|^2}{n^2} \cdot \sum_{i=1}^n (i^2 \cdot S_i^2)\right) \leq \mathcal{O}(|U|^2 \cdot |AT|^2).$$

Algorithm 2: Attribute Reduction Based on LIFT and Random Sampling

Input: Decision system \mathcal{DS} , constraint C_ρ^U , amount of sample groups n .
Output: One reduct A .

- 1 Create the positive class samples P_k and negative class samples N_k by, respectively, using the Equations (8) and (9);
- 2 Utilize k -means clustering over P_k and N_k , get every cluster centers according to Equation (10);
- 3 Collect all the cluster centers into U_1 ;
- 4 Randomly divide other sample into $n - 1$ groups with same size: U_2, \dots, U_n ;
- 5 Get reduct A_1 over U_1 by using Algorithm 1;
- 6 **For** $i = 2 : n$ **do**
- 7 // i represents the iteration quantity
- 8 $A_i = A_{i-1}$;
- 9 // The guidance of the selection of attribute by previous reduct
- 10 **While** $C_\rho^{\cup_{j=1}^i U_j}$ is not met **do**
- 11 // Determine the required attributes
- 12 $\forall a \in AT - A_i$, perform information granulation over $\cup_{j=1}^i U_j$ by using $A_i \cup \{a\}$;
- 13 Evaluate a by calculating $\rho(\cup_{j=1}^i U_j, A_i \cup \{a\})$;
- 14 Choose an appropriate attribute $b \in AT - A_i$ by the standard;
- 15 $A_i = A_i \cup \{b\}$;
- 16 Calculate $\rho(\cup_{j=1}^i U_j, A_i)$;
- 17 **End**
- 18 **End**
- 19 $A = A_n$;
- 20 **Repeat**
- 21 // Remove the redundant attributes
- 22 $\forall c \in A$, perform information granulation over U by using $A - \{c\}$;
- 23 Compute $\rho(U, A - \{c\})$;
- 24 **If** C_ρ^U is met
- 25 | $A = A - \{c\}$;
- 26 **End**
- 27 **Until** A remains unchanged;
- 28 **Return** A .

In this section, according to the above strategy, an attribute reduction algorithm called ARLRS is proposed as Algorithm 2. In ARLRS, LIFT algorithm is used to choose the samples from the dataset as the members of the first group, and random sampling strategy is used to divide the rest of samples into groups. Generally, the reduct of the first sample group is calculated quickly by Algorithm 1. For every subsequent sample group, if the given constrain is satisfied, the attribute reduction calculation of current sample group can be skipped; if the given constrain is not satisfied, the previous sample groups are integrated into the current sample group, then attribute reduction calculation of this integrated group is carried out. As a result, by gradually integrating many previous sample groups, the classification performance of the temporary reduct over the current sample group will be improved until the attribute reduction constrain is met.

4. Analysis of Experiment

4.1. Datasets

To illustrate the effectiveness of the proposed ARLRS algorithm, 17 UCI datasets were chosen for the tests. The UCI datasets is a collection of databases, top level theory, and data generators used by the machine learning to conduct empirical evaluations of machine learning algorithms. David Aha and his fellow graduate students at UC Irvine established the archive as an ftp archive in 1987. Since then, it has been widely used as a key source of machine learning datasets by students, instructors, and researchers all over the world. All the datasets this paper used were download on: <https://archive.ics.uci.edu/ml/datasets>.

php (accessed on 12 January 2022). The Table 1 contains full descriptions of 17 datasets. All experiment were conducted on a personal computer running Windows 10, with an Intel Core i7-6700HQ CPU (2.60 GHz) and 16.00 GB of RAM. Matlab R2019b is the programming language. In order to further analyze the advantages of ARLRS algorithm, the experiment will be shown in two aspects: time consumption and classification performance.

Table 1. Datasets used in the experiment.

ID	Datasets	# Samples	# Attributes	# Labels	Attributes Type
1	Australian Credit Approval [36]	690	14	2	Real
2	Breast Cancer Wisconsin (Diagnostic) [37]	569	30	2	Real
3	Cardiotocography [38]	2126	21	10	Real
4	Connectionist Bench (Sonar, Mines vs. Rocks) [39]	208	60	2	Real
5	Crowdsourced Mapping [40]	10,845	28	6	Real
6	Diabetic Retinopathy Debrecen [41]	1151	19	2	Integer & Real
7	DrivFace [42]	606	6400	3	Real
8	Forest Type Mapping [43]	523	27	4	Integer & Real
9	Glass Identification [44]	214	9	6	Real
10	Ionosphere [45]	351	34	2	Integer & Real
11	MAGIC Gamma Telescope [46]	19,020	28	6	Real
12	Parkinson Multiple Sound Recording [47]	1208	26	2	Real
13	QSAR Biodegradation [48]	1055	41	2	Real
14	Musk (Version 1) [49]	476	166	2	Integer
15	Page Blocks Classification [50]	5473	10	5	Integer & Real
16	Urban Land Cover [51]	675	147	9	Integer & Real
17	Quality Assessment of Digital Colposcopies [52]	287	62	2	Real

4.2. Basic Experiment Setting

In this paper, the neighborhood rough set [53–58] was used as the model for deriving reducts. It should be noted that the computational results of the neighborhood rough set are heavily reliant on the radius chosen, and we pick 0.2 as our radius to illustrate the generality of ARLRS algorithm. Moreover, the concept of approximation quality (Equation (6)) is selected as the constrain of ARLRS algorithm. The approximation quality describes the ability of the approximation in terms of attribute reduction. In general, the higher the degree of approximation quality is, the better performance the selected attribute has.

Five classical algorithms are chosen to compare with the proposed ARLRS algorithm. The four algorithms of them have been introduced detailed in previous context of this paper, that are FGSAR [59–61], BBSAR [10], PAAR [11] and RSAR [12]. The fifth compared algorithm is the Attribute Group for Attribute Reduction (AGAR) [62]. AGAR is also a sample based algorithm and has excellent time efficiency and classification performance. Adding AGAR to the comparison can better help us test the performance of ARLRS algorithm.

4.3. Time Consumption

In this subsection, the time consumption of five compared algorithms (FGSAR, BBSAR, PAAR, RSAR and AGAR) and ARLRS algorithm on computing reducts are listed in Table 2. For every dataset, the best time consumption is bold.

Table 2 shows that ARLRS performs admirably on time consumption compared to the other five classical algorithms. Using the “QSAR Biodegradation” (ID 13) dataset as an example, the time consumptions of RSAR, FGSAR, PAAR, BBSAR and AGAR are 11.9454, 13.6468, 13.4522, 12.5647 and 12.3949 seconds, respectively. The proposed ARLRS algorithm only takes 9.6467 seconds. Obviously, ARLRS algorithm has the ability to remarkable accelerate the process of computing reduct.

Table 2. The time consumption (unit is second) of six attribute reduction algorithms.

ID	ARLRS	RSAR	FGSAR	PAAR	BBSAR	AGAR
1	1.3439	1.6046	1.9683	1.9533	3.0378	2.0001
2	2.8514	3.2616	3.7553	3.6474	4.9878	6.3496
3	1.7156	2.3740	3.6215	3.3452	4.0248	2.6923
4	3.114	3.8712	4.6534	4.4532	4.9458	6.3526
5	0.0331	0.0351	0.3561	0.3424	0.4138	0.0493
6	1.5043	1.6542	3.3462	3.3977	2.7745	3.4527
7	21.5071	21.6074	22.9543	22.6464	32.7441	24.2342
8	3.2707	4.9173	10.2732	10.2480	4.4253	4.5429
9	36.6825	39.9173	55.4524	54.2809	46.355	40.6939
10	2.2156	2.3740	3.5643	3.5754	4.0248	2.5551
11	32.6542	34.7645	35.3475	35.3342	35.4578	34.9578
12	2.5127	2.7545	3.1642	3.0008	2.8361	3.147
13	9.6467	11.9454	13.6468	13.4522	12.5647	12.3949
14	16.9655	18.3361	19.7468	19.2312	19.1485	18.5435
15	44.4554	46.1265	48.4641	48.8529	45.5659	48.4844
16	3.6125	4.0415	4.8353	4.1237	4.7264	4.9645
17	38.4124	41.5415	43.8515	43.4529	45.4554	44.5455

To further demonstrate the superiority of ARLRS algorithm, the speedup ratios of ARLRS algorithm against the five compared algorithms (FGSAR, BBSAR, PAAR, RSAR and AGAR) are shown in Table 3 and the lengths of reducts produced by overall six algorithms are shown in Table 4. It is worth mentioning that FGSAR and PAAR have similar mechanism, so their reduct lengths have symmetry. For every dataset, the shortest length of reducts produced by six algorithms is set bold.

Table 3. The speed up ratio of five compared algorithms to ARLRS.

ID	RSAR	FGSAR	PAAR	BBSAR	AGAR
1	16.37	31.19	41.17	30.34	32.81
2	12.58	24.07	21.82	20.53	19.31
3	27.73	35.12	27.82	18.21	26.34
4	19.63	33.14	30.13	25.75	7.22
5	2.36	23.38	2.65	2.71	17.78
6	9.06	14.64	42.91	34.52	27.98
7	13.15	19.01	18.62	4.57	5.99
8	8.11	33.85	32.42	34.91	9.86
9	20.13	37.84	38.03	24.24	13.29
10	10.23	11.09	22.97	23.13	30.23
11	6.07	7.02	7.88	7.91	6.59
12	25.97	20.59	16.27	11.40	20.16
13	19.24	9.39	7.71	23.22	22.17
14	25.09	18.28	13.71	12.93	6.77
15	3.62	8.27	6.00	2.44	8.31
16	15.23	13.88	14.76	17.84	20.76
17	31.25	4.33	6.12	7.32	15.9
avg	15.63	20.29	20.64	17.79	17.14

According to Table 3, all speedup ratios of ARLRS algorithm against other five classic sample based algorithms are substantially greater than 1. Even that average speedup ratio approaches 15. These findings show that the proposed ARLRS algorithm can speed up the process of computing reducts from datasets. Furthermore, Table 4 shows that in the most of datasets, the lengths of reducts generated by ARLRS are substantially shorter than those acquired by the other five algorithms. The reducts with shorter length often indicate that ARLRS needs fewer iterations to get the reducts with satisfying classification performance.

From this viewpoint, it also shows that ARLRS can effectively reduce the time consumption of attribute reduction.

Table 4. Reduct lengths of six algorithms.

ID	ARLRS	RSAR	FGSAR	PAAR	BBSAR	AGAR
1	7.14	7.32	8.03	8.03	8.03	7.17
2	8.67	8.21	9.19	9.19	9.19	8.11
3	7.38	8.66	7.57	7.57	7.57	8.45
4	7.13	7.43	8.17	8.17	8.17	7.43
5	14.34	15.32	14.38	14.38	14.38	16.64
6	9.46	9.21	8.32	8.32	8.32	8.76
7	8.74	9.87	7.15	7.15	7.15	9.75
8	8.65	9.07	9.25	9.25	9.25	8.98
9	7.91	9.87	8.57	8.57	8.57	8.07
10	8.65	8.97	8.78	8.78	8.78	9.21
11	23.87	24.58	25.32	25.32	25.32	27.98
12	7.03	10.34	10.01	10.01	10.01	9.28
13	6.98	9.01	8.34	8.34	8.34	7.32
14	8.75	8.65	8.90	8.90	8.90	10.76
15	13.66	15.32	16.24	16.24	16.24	15.23
16	14.31	15.09	15.65	15.65	15.65	16.32
17	8.02	8.06	8.19	8.19	8.19	9.15

Among the above six attribute reduction algorithms, ARLRS and RSAR are designed both based on random sampling. It is not difficult to find that for the two algorithms with the similar mechanism, ARLRS algorithm has obvious advantage on the time consumption. Especially on high-dimensional datasets such as “Musk (Version 1)” (ID 14), “DrivFace” (ID 7), “Quality Assessment of Digital Colposcopies” (ID 17) and “MAGIC Gamma Telescope” (ID 11), ARLRS algorithm performs particularly well. In order to better show the advantage of ARLRS algorithm on high-dimensional datasets, Figure 1 shows the time consumption over these four datasets. Obviously, the time consumption of ARLRS is less than that of RSAR over the four high-dimensional datasets.

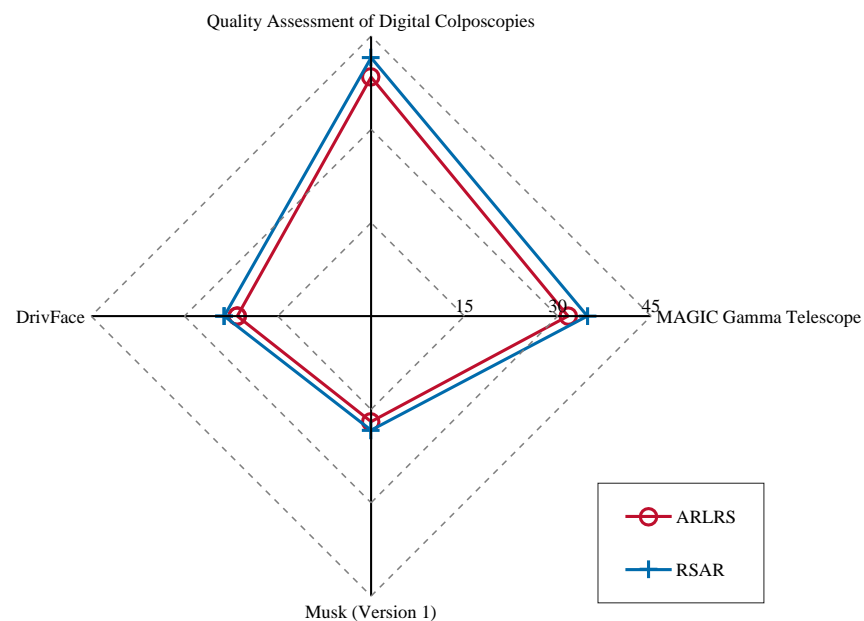


Figure 1. Time consumption of attribute reduction on high-dimensional datasets.

It is worth mentioning that ARLRS will filter out the samples with low representation, then only evaluate the samples with most representation. The filtration process leads to reduce greatly the time consumption of samples evaluation in the process of attribute reduction. Therefore, comparing with other attribute reduction strategies, the proposed ARLRS algorithm has better performance in time consumption.

4.4. Classification Performance

K-Nearest Neighbor (KNN, K parameter will be set as 3) and Support Vector Machine (SVM) classifiers are used to test the classification accuracies of different reducts produced by six algorithms. The average values of the classification accuracies of generated reducts are provided in Tables 5 and 6. It is worth mentioning that FGSAR and PAAR have similar mechanism, so the classification performance of them have symmetry. For every dataset, the best classification accuracy is bold.

According to Tables 5 and 6, ARLRS's classification performance is somewhat greater than the other five algorithms. Especially for high-dimensional dataset "DrivFace" (ID 7), the classification accuracies of proposed ARLRS are 0.9089 on KNN classifier and 0.8841 on SVM classifier, respectively, which perform better than other five algorithms. As a result, regardless of which two classifiers are utilized, the reducts obtained from ARLRS deliver satisfying classification performance.

It is because LIFT can filter out samples with low representation, some candidate attributes will be deleted after the attribute reduction of the first sample group. The subsequent reduction process can receive better guidance for selecting attributes with better classification performance. This mechanism enables the attribute reduction process to focus on attributes with most representation. This is why the classification performances of reducts produced by ARLRS can be satisfying.

Table 5. Based on KNN classifier, the classification accuracies of reducts produced by six attribute reduction algorithms.

ID	ARLRS	RSAR	FGSAR	PAAR	BBSAR	AGAR
1	0.8555	0.8742	0.8168	0.8168	0.8557	0.9175
2	0.9785	0.9000	0.8534	0.8534	0.9347	0.9456
3	0.9399	0.9320	0.8824	0.8824	0.8356	0.8943
4	0.9569	0.9642	0.8074	0.8074	0.9741	0.9456
5	0.8848	0.8351	0.8022	0.8022	0.8149	0.8624
6	0.8449	0.8272	0.9074	0.9074	0.8736	0.8488
7	0.9089	0.9003	0.8075	0.8075	0.8572	0.8197
8	0.8437	0.8326	0.8731	0.8731	0.8186	0.8279
9	0.8441	0.8741	0.8507	0.8507	0.9117	0.7986
10	0.9575	0.9320	0.9269	0.9269	0.8356	0.8943
11	0.9145	0.8487	0.8366	0.8366	0.8869	0.8747
12	0.9035	0.9055	0.9266	0.9266	0.8845	0.8647
13	0.8364	0.8641	0.8644	0.8644	0.9153	0.8279
14	0.9634	0.9534	0.9128	0.9128	0.8934	0.8634
15	0.9314	0.9169	0.8674	0.8674	0.8467	0.8469
16	0.9299	0.9064	0.8796	0.8796	0.9074	0.8534
17	0.8036	0.8712	0.8642	0.8642	0.8541	0.8779

Table 6. Based on SVM classifier, the classification accuracies of reducts produced by six attribute reduction algorithms.

ID	ARLRS	RSAR	FGSAR	PAAR	BBSAR	AGAR
1	0.8396	0.8614	0.8318	0.8318	0.8496	0.8779
2	0.8984	0.8401	0.8302	0.8302	0.7121	0.7326
3	0.8351	0.8741	0.9164	0.9164	0.8813	0.8801
4	0.8433	0.8381	0.8041	0.8041	0.8701	0.8912
5	0.9103	0.8701	0.9328	0.9328	0.9488	0.8468
6	0.9434	0.9362	0.9723	0.9723	0.9356	0.9147
7	0.8841	0.8766	0.8452	0.8452	0.8141	0.7998
8	0.8321	0.8311	0.9723	0.9723	0.7318	0.7323
9	0.9194	0.9248	0.9049	0.9049	0.8722	0.8413
10	0.8954	0.8888	0.7763	0.7763	0.8052	0.7113
11	0.9165	0.8971	0.9460	0.9460	0.8741	0.8492
12	0.9365	0.9207	0.9194	0.9194	0.8279	0.7812
13	0.9153	0.8803	0.7436	0.7436	0.7799	0.8940
14	0.8915	0.8766	0.8622	0.8622	0.8348	0.8786
15	0.9255	0.8903	0.8802	0.8802	0.8584	0.9205
16	0.9044	0.8909	0.8633	0.8633	0.8261	0.8759
17	0.8759	0.9204	0.8816	0.8816	0.9362	0.8294

4.5. Discussion

In the above experimental analysis, it can be concluded that ARLRS algorithm often performs well in big data, both in terms of time consumption and classification performance, for this conclusion, the following reasons can be given.

- (1) The higher the dimension of the datasets is, the more information the selected sample can carry. Then, the more information can better help to guide the subsequent attribute reduction. Therefore, ARLRS algorithm performs well in handling big data.
- (2) Due to the low value density of big data, it is often necessary to preferentially extract relevant and useful information from a large amount of datasets. The above-mentioned needs are well solved by introducing LIFT. The most representative samples found by LIFT make the unclear information under each label have available structure, from the perspective of samples, thus different samples form a certain association. Such preprocessing is very effective in the processing big data, so ARLRS will have certain advantages in processing big data.

5. Conclusions and Future Perspectives

This paper combines LIFT with random sampling strategy to proposed an attribute reduction algorithm call ARLRS. The utilization of LIFT reduces the impact of overfitting in the first sample group, which can produce the reduct being able to provide better guidance for subsequent attribute reduction. The retainment of random sampling strategy is helpful to reduce the impact of sample distribution. Due to the fact that LIFT can mine the samples with most representation, the proposed ARLRS algorithm also has satisfying results in classification performance. The experimental results over 17 UCI datasets show that the above conclusions are effective and credible.

The following research directions will be of great value as our next work.

- (1) In light of the uncertainty existed in constraints and classification performances, the resultant reduct may result in over-fitting. As a result, in future investigation, we will try balance the efficiency of attribute reduction with the classification performance.
- (2) The attribute reduction strategy presented in this research is only applied from the sample's perspective. Therefore, we will try to investigate some novel algorithms taking into account both samples and attributes for improving more efficiency.

Author Contributions: Conceptualization, T.X.; methodology, T.X.; software, J.C.; validation, T.X.; formal analysis, J.C.; investigation, J.C.; resources, J.C.; data curation, Q.C.; writing—original draft preparation, Q.C.; writing—review and editing, Q.C.; visualization, Q.C.; supervision, T.X.; project administration, T.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (No. 62006099), the Natural Science Foundation of Jiangsu Provincial Colleges and Universities (20KJB520010), and the Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province (No. OBDMA202104).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Chen, D.G.; Yang, Y.Y.; Dong, Z. An incremental algorithm for attribute reduction with variable precision rough sets. *Appl. Soft Comput.* **2016**, *45*, 129–149. [[CrossRef](#)]
2. Jiang, Z.H.; Yang, X.B.; Yu, H.L.; Liu, D.; Wang, P.X.; Qian, Y.H. Accelerator for multi-granularity attribute reduction. *Knowl.-Based Syst.* **2019**, *177*, 145–158. [[CrossRef](#)]
3. Ju, H.R.; Yang, X.B.; Yu, H.L.; Li, T.J.; Yu, D.J.; Yang, J.Y. Cost-sensitive rough set approach. *Inf. Sci.* **2016**, *355–356*, 282–298. [[CrossRef](#)]
4. Qian, Y.H.; Wang, Q.; Cheng, H.H.; Liang, J.Y.; Dang, C.Y. Fuzzy-rough feature selection accelerator. *Fuzzy Sets Syst.* **2015**, *258*, 61–78. [[CrossRef](#)]
5. Wei, W.; Cui, J.B.; Liang, J.Y.; Wang, J.H. Fuzzy rough approximations for set-valued data. *Inf. Sci.* **2016**, *360*, 181–201. [[CrossRef](#)]
6. Wei, W.; Wu, X.Y.; Liang, J.Y.; Cui, J.B.; Sun, Y.J. Discernibility matrix based incremental attribute reduction for dynamic data. *Knowl.-Based Syst.* **2018**, *140*, 142–157. [[CrossRef](#)]
7. Dong, L.J.; Chen, D.G. Incremental attribute reduction with rough set for dynamic datasets with simultaneously increasing samples and attributes. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 213–227. [[CrossRef](#)]
8. Zhang, A.; Chen, Y.; Chen, L.; Chen, G.T. On the NP-hardness of scheduling with time restrictions. *Discret. Optim.* **2017**, *28*, 54–62. [[CrossRef](#)]
9. Guan, L.H. A heuristic algorithm of attribute reduction in incomplete ordered decision systems. *J. Intell. Fuzzy Syst.* **2019**, *36*, 3891–3901. [[CrossRef](#)]
10. Liu, Y.; Huang, W.L.; Jiang, Y.L.; Zeng, Z.Y. Quick attribute reduct algorithm for neighborhood rough set model. *Inf. Sci.* **2014**, *271*, 65–81. [[CrossRef](#)]
11. Qian, Y.H.; Liang, J.Y.; Pedrycz, W.; Dang, C.Y. Positive approximation: An accelerator for attribute reduction in rough set theory. *Artif. Intell.* **2010**, *174*, 597–618. [[CrossRef](#)]
12. Chen, Z.; Liu, K.Y.; Yang, X.B.; Fujita, H. Random sampling accelerator for attribute reduction. *Int. J. Approx. Reason.* **2022**, *140*, 75–91. [[CrossRef](#)]
13. Wang, K.; Thrampoulidis, C. Binary classification of gaussian mixtures: Abundance of support vectors, benign overfitting, and regularization. *SIAM J. Math. Data Sci.* **2022**, *4*, 260–284. [[CrossRef](#)]
14. Bejani, M.M.; Ghatee, M.A. A systematic review on overfitting control in shallow and deep neural networks. *Artif. Intell. Rev.* **2021**, *54*, 6391–6438. [[CrossRef](#)]
15. Park, Y.; Ho, J.C. Tackling overfitting in boosting for noisy healthcare data. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 2995–3006. [[CrossRef](#)]
16. Zhang, M.L.; Wu, L. LIFT: Multi-label learning with label-specific features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 107–120. [[CrossRef](#)] [[PubMed](#)]
17. Hu, Q.H.; Zhang, L.; Chen, D.G.; Pedrycz, W.; Yu, D. Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications. *Int. J. Approx. Reason.* **2010**, *51*, 453–471. [[CrossRef](#)]
18. Sun, L.; Wang, L.Y.; Ding, W.P.; Qian, Y.H.; Xu, J.C. Feature selection using fuzzy neighborhood entropy-nased uncertainty measures for fuzzy neighborhood mltigranulation rough sets. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 19–33. [[CrossRef](#)]
19. Wang, Z.C.; Hu, Q.H.; Qian, Y.H.; Chen, D.E. Label enhancement-based feature selection via fuzzy neighborhood discrimination index. *Knowl.-Based Syst.* **2022**, *250*, 109119. [[CrossRef](#)]
20. Li, W.T.; Wei, Y.L.; Xu, W.H. General expression of knowledge granularity based on a fuzzy relation matrix. *Fuzzy Sets Syst.* **2022**, *440*, 149–163. [[CrossRef](#)]
21. Liu, F.L.; Zhang, B.W.; Ciucci, D.; Wu, W.Z.; Min, F. A comparison study of similarity measures for covering-based neighborhood classifiers. *Inf. Sci.* **2018**, *448–449*, 1–17. [[CrossRef](#)]

22. Ma, X.A.; Yao, Y.Y. Min-max attribute-object bireducts: On unifying models of reducts in rough set theory. *Inf. Sci.* **2019**, *501*, 68–83. [CrossRef]
23. Xu, T.H.; Wang, G.Y.; Yang, J. Finding strongly connected components of simple digraphs based on granulation strategy. *Int. J. Approx. Reason.* **2020**, *118*, 64–78. [CrossRef]
24. Jia, X.Y.; Rao, Y.; Shang, L.; Li, T.J. Similarity-based attribute reduction in rough set theory: A clustering perspective. *Int. J. Approx. Reason.* **2020**, *11*, 1047–1060. [CrossRef]
25. Ding, W.P.; Pedrycz, W.; Triguero, I.; Cao, Z.H.; Lin, C.T. Multigranulation supertrust model for attribute reduction. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 1395–1408. [CrossRef]
26. Chu, X.L.; Sun, B.D.; Chu, X.D.; Wu, J.Q.; Han, K.Y.; Zhang, Y.; Huang, Q.C. Multi-granularity gominance rough concept attribute reduction over hybrid information systems and its application in clinical decision-making. *Inf. Sci.* **2022**, *597*, 274–299. [CrossRef]
27. Yuan, Z.; Chen, H.M.; Yang, X.L.; Li, T.R.; Liu, K.Y. Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction. *Knowl.-Based Syst.* **2021**, *231*, 107398. [CrossRef]
28. Zhang, Q.Y.; Chen, Y.; Zhang, G.Q.; Li, Z.; Chen, L.; Wen, C.F. New uncertainty measurement for categorical data based on fuzzy information structures: An application in attribute reduction. *Inf. Sci.* **2021**, *580*, 541–577. [CrossRef]
29. Ding, W.P.; Wang, J.D.; Wang, J.H. Multigranulation consensus fuzzy-rough based attribute reduction. *Knowl.-Based Syst.* **2020**, *198*, 105945. [CrossRef]
30. Chen, Y.; Yang, X.B.; Li, J.H.; Wang, P.X.; Qian, Y.H. Fusing attribute reduction accelerators. *Inf. Sci.* **2022**, *587*, 354–370. [CrossRef]
31. Yan, W.W.; Ba, J.; Xu, T.H.; Yu, H.L.; Shi, J.L.; Han, B. Beam-influenced attribute selector for producing stable reduct. *Mathematics* **2022**, *10*, 533. [CrossRef]
32. Ganguly, S.; Bhowal, P.; Oliva, D.; Sarkar, R. BLeafNet: A bonferroni mean operator based fusion of CNN models for plant identification using leaf image classification. *Ecol. Inform.* **2022**, *69*, 101585. [CrossRef]
33. Zhang, C.F.; Feng, Z.L. Convolutional analysis operator learning for multifocus image fusion. *Signal Process. Image Commun.* **2022**, *103*, 116632. [CrossRef]
34. Jiang, Z.H.; Dou, H.L.; Song, J.J.; Wang, P.X.; Yang, X.B.; Qian, Y.H. Data-guided multi-granularity selector for attribute reduction. *Artif. Intell. Rev.* **2021**, *51*, 876–888. [CrossRef]
35. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.D.; Silverman, R.; Wu, A.Y. An efficient k -means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [CrossRef]
36. Quinlan, J.R. Simplifying decision trees. *Int. J.-Hum.-Comput. Stud.* **1999**, *51*, 497–510. [CrossRef]
37. Street, N.; Wolberg, W.; Mangasarian, O. Nuclear feature extraction for breast tumor diagnosis. *Int. Symp. Electron. Imaging Sci. Technol.* **1999**, *1993*, 861–870. [CrossRef]
38. Ayres-de-Campos, D.; Bernardes, J.; Garrido, A.; Marques-de-sá, J.; Pereira-leite, L. SisPorto 2.0: A program for automated analysis of cardiocograms. *J. Matern.-Fetal Med.* **2000**, *9*, 311–318. [CrossRef]
39. Gorman, R.P.; Sejnowski, T.J. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Netw.* **1988**, *16*, 75–89. [CrossRef]
40. Johnson, B.A.; Iizuka, K. Integrating open street map crowd sourced data and landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the laguna de bay area of the philippines. *Appl. Geogr.* **2016**, *67*, 140–149. [CrossRef]
41. Antal, B.; Hajdu, A. An ensemble-based system for automatic screening of diabetic retinopathy. *Knowl.-Based Syst.* **2014**, *60*, 20–27. [CrossRef]
42. Diaz-Chito, K.; Hernández, A.; López, A. A reduced feature set for driver head pose estimation. *Appl. Soft Comput.* **2016**, *45*, 98–107. [CrossRef]
43. Johnson, B.; Tateishi, R.; Xie, Z. Using geographically-weighted variables for image classification. *Remote Sens. Lett.* **2012**, *3*, 491–499. [CrossRef]
44. Evett, I.W.; Spiehler, E.J. Rule induction in forensic science. *Knowl. Based Syst.* **1989**, 152–160. Available online: <https://doi.org/10.5555/67040.67055> (accessed on 15 August 2022). [CrossRef]
45. Sigillito, V.; Wing, S.; Hutton, L.; Baker, K. Classification of radar returns from the ionosphere using neural networks. *Johns Hopkins APL Tech. Dig.* **1989**, *10*, 876–890. [CrossRef]
46. Bock, R.K.; Chilingarian, A.; Gaug, M.; Hakl, F.; Hengstebeck, T.; Jiřina, M.; Klaschka, J.; Kotrč, E.; Savický, P.; Towers, S.; et al. Methods for multidimensional event classification: A case study using images from a cherenkov gamma-ray telescope. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **2004**, *516*, 511–528. [CrossRef]
47. Sakar, B.E.; Isenkul, M.E.; Sakar, C.O.; Sertbas, A.; Gurgun, F.; Delil, S.; Apaydin, H.; Kursun, O. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 828–834. [CrossRef]
48. Mansouri, K.; Ringsted, T.; Ballabio, D.; Todeschini, R.; Consonni, V. Quantitative structure–activity relationship models for ready biodegradability of chemicals. *J. Chem. Inf. Model.* **2013**, *53*, 867–878. [CrossRef]
49. Dietterich, T.G.; Lathrop, R.H.; Lozano-Pérez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.* **1997**, *89*, 31–71. [CrossRef]
50. Malerba, D.; Esposito, F.; Semeraro, G. A further comparison of simplification methods for decision-tree induction. In *Learning from Data*; Springer: New York, NY, USA, 1996; pp. 365–374. [CrossRef]
51. Johnson, B.; Xie, Z. Classifying a high resolution image of an urban area using super-object information. *ISPRS J. Photogramm. Remote Sens.* **2013**, *83*, 40–49. [CrossRef]

52. Fernandes, K.; Cardoso, J.S.; Fernandes, J. Transfer learning with partial observability applied to cervical cancer screening. *Pattern Recognit. Image Anal.* **2017**, *243*–250. [[CrossRef](#)]
53. Sun, L.; Wang, T.X.; Ding, W.P.; Xu, J.C.; Lin, Y.J. Feature selection using fisher score and multilabel neighborhood rough sets for multilabel classification. *Inf. Sci.* **2021**, *578*, 887–912. [[CrossRef](#)]
54. Luo, S.; Miao, D.Q.; Zhang, Z.F.; Zhang, Y.J.; Hu, S.D. A neighborhood rough set model with nominal metric embedding. *Inf. Sci.* **2020**, *520*, 373–388. [[CrossRef](#)]
55. Chu, X.L.; Sun, B.Z.; Li, X.; Han, K.Y.; Wu, J.Q.; Zhang, Y.; Huang, Q.C. Neighborhood rough set-based three-way clustering considering attribute correlations: An approach to classification of potential gout groups. *Inf. Sci.* **2020**, *535*, 28–41. [[CrossRef](#)]
56. Shu, W.H.; Qian, W.B.; Xie, Y.H. Incremental feature selection for dynamic hybrid data using neighborhood rough set. *Knowl.-Based Syst.* **2020**, *194*, 105516. [[CrossRef](#)]
57. Wan, J.H.; Chen, H.M.; Yuan, Z.; Li, T.R.; Sang, B.B. A novel hybrid feature selection method considering feature interaction in neighborhood rough set. *Knowl.-Based Syst.* **2021**, *227*, 107167. [[CrossRef](#)]
58. Sang, B.B.; Chen, H.M.; Yang, L.; Li, T.R.; Xu, W.H.; Luo, W.H. Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set. *Knowl.-Based Syst.* **2021**, *227*, 107223. [[CrossRef](#)]
59. Hu, Q.H.; Xie, Z.X.; Yu, D. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognit.* **2007**, *40*, 3509–3521. [[CrossRef](#)]
60. Jensen, R.; Qiang, S. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets Syst.* **2004**, *141*, 469–485. [[CrossRef](#)]
61. Xu, S.P.; Yang, X.B.; Yu, H.L.; Yu, D.J.; Yang, J.Y.; Tsang, E.C. Multi-label learning with label-specific feature reduction. *Knowl.-Based Syst.* **2016**, *104*, 52–61. [[CrossRef](#)]
62. Chen, Y.; Liu, K.Y.; Song, J.J.; Yang, X.B.; Qian, Y.H. Attribute group for attribute reduction. *Inf. Sci.* **2020**, *535*, 64–80. [[CrossRef](#)]