Intelligent Automation & Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.037874
Article

Fusing Supervised and Unsupervised Measures for Attribute Reduction

Tianshun Xing, Jianjun Chen*, Taihua Xu and Yan Fan

School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, 212100, China
*Corresponding Author: Jianjun Chen. Email: jianjunchen@just.edu.cn
Received: 19 November 2022; Accepted: 06 February 2023

Abstract: It is well-known that attribute reduction is a crucial action of rough set. The significant characteristic of attribute reduction is that it can reduce the dimensions of data with clear semantic explanations. Normally, the learning performance of attributes in derived reduct is much more crucial. Since related measures of rough set dominate the whole process of identifying qualified attributes and deriving reduct, those measures may have a direct impact on the performance of selected attributes in reduct. However, most previous researches about attribute reduction take measures related to either supervised perspective or unsupervised perspective, which are insufficient to identify attributes with superior learning performance, such as stability and accuracy. In order to improve the classification stability and classification accuracy of reduct, in this paper, a novel measure is proposed based on the fusion of supervised and unsupervised perspectives: (1) in terms of supervised perspective, approximation quality is helpful in quantitatively characterizing the relationship between attributes and labels; (2) in terms of unsupervised perspective, conditional entropy is helpful in quantitatively describing the internal structure of data itself. In order to prove the effectiveness of the proposed measure, 18 University of CaliforniaIrvine (UCI) datasets and 2 Yale face datasets have been employed in the comparative experiments. Finally, the experimental results show that the proposed measure does well in selecting attributes which can provide distinguished classification stabilities and classification accuracies.

Keywords: Approximation quality; attribute reduction; conditional entropy; neighborhood rough set

1 Introduction

In the era of big data, a large number of irrelevant and redundant attributes are usually generated in practical applications, which will bring a series of problems in data processing, such as over-fitting, high computing cost and insufficient classification performance. Attribute reduction [1–4] is one of the effective methods to deal with this problem in rough set theory [5]. As a process of deleting irrelevant attributes and preserving key attributes, attribute reduction can reduce computational complexity and time by eliminating the influence of irrelevant attributes and noise. The result of attribute reduction

is called reduct. Reduct is a new data which is composed of key attributes. In certain criterion of evaluation, such reduct can achieve or go beyond the performance of raw data [6]. As a dimension reduction method, the attribute reduction based on rough set theory consists of three stages, which are data representation under rough set, construction of attribute evaluation constraints and selection of search strategy. In the first stage, raw data can be described as a triple set, they are samples, conditional attributes and decision attributes. In second stage, the constraint can be constructed by using some related measures of rough set in certain learning perspective. There are many related measures in rough set theory, such as approximation quality [7], conditional entropy [8], and decision error rate [9]. And the learning perspective can be usually divided into two parts, supervised learning and unsupervised learning. Different measures correspond to different constraints [10]. In the third stage, using suitable search strategy can obtain the reduct effectively. Serving as a criterion of evaluation, constraint controls the condition of terminating search. Naturally, it may be concluded that different constraints imply different reducts, and then different learning performances.

With a literature review, it is noticed that most studies about constraint based on either supervised perspective or unsupervised perspective have been extensively explored [11]. In other words, the measure based on only one perspective is used to form a constraint of attribute reduction, and then the qualified reduct can be sought out through such a constraint. For example, Jiang et al. [12] and Yuan et al. [13] have investigated attribute reductions with respect to supervised information granulation and the relative supervised measures, respectively; Yang et al. [14] have introduced a measure, called fuzzy complementary entropy, into the attribute reduction based on the unsupervised framework. Nevertheless, either supervised measure or unsupervised measure is considered, the measure based on single perspective may have some limitations as follows.

(1) The measure based on single perspective neglects the diverse evaluations [15,16]. That is, it may not identify more confident attributes. The potential reason is that if a measure is fixed to perform attribute reduction, then within the iterations in searching reduct, only such a measure can assess the importance of candidate attributes, so it follows that immediate results of the evaluation may be invalidated if some other measures are further required.
(2) The measure based on single perspective ignores the complex constraint [17]. It may not effectively terminate the process of attribute selection. For example, supposed that conditional entropy serves as a measure to evaluate attributes [18,19], a derived reduct is only equipped with the single characteristic required by such an evaluation, no other types of uncertainty characteristics and learning abilities have been fully considered.

To overcome the limitations mentioned above, a new measure is proposed, which is a fusion of measures defined in both supervised and unsupervised perspectives. Fig. 1 shows the flowchart of attribute reduction, the dotted box presents our main work. Approximation quality and conditional entropy are classical measures and have been widely accepted in the field of rough set. The former can be used to build a bridge between attributes and labels [20], while the latter can quantitatively characterize the uncertainty of the data itself [21]. In our study, approximation quality from a supervised perspective and conditional entropy from an unsupervised perspective are employed to fuse into a new measure. Through abundant comparative experiments, the superiorities of new measure can be summarized as: (1) it is equipped with the distinguishing characteristic of multiple perspectives [22–24]; (2) it has more complex constraints than measure based on single perspective [25] and then more appropriate attributes are selected [26]; (3) it can not only quantitatively describe the relationship between attributes and label [27], but also quantitatively describe the internal structure of data itself [28,29].

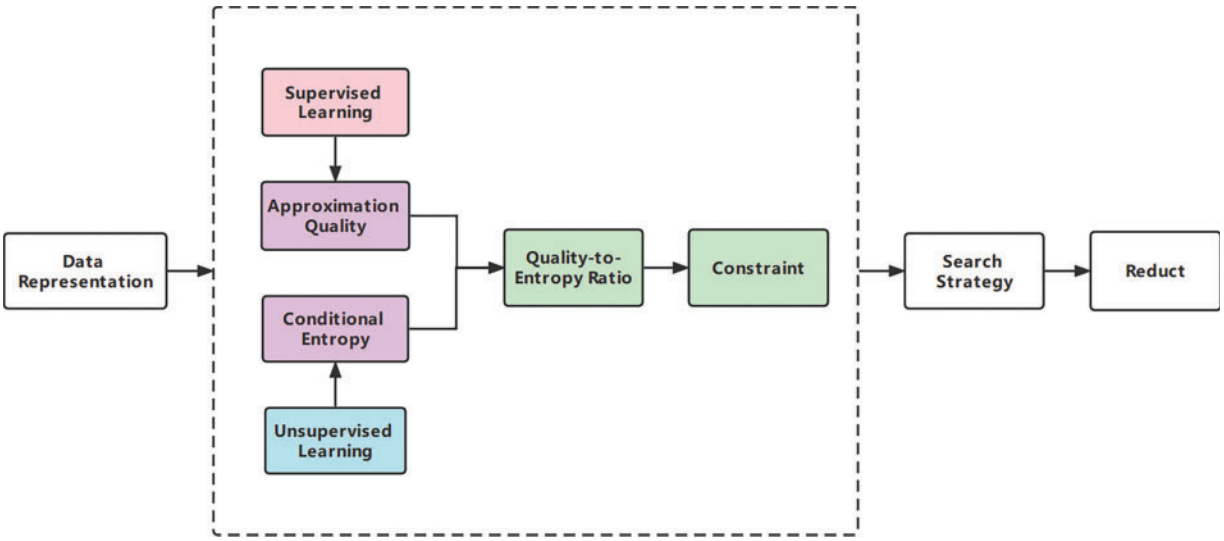

Figure 1: Fusing supervised and unsupervised measures for attribute reduction

The rest of this paper is organized as follows: Section 2 introduces the basic notions about the rough set, related measures, and attribute reduction. The proposed novel measure and searching process of solving reduct are elaborated in Section 3. Section 4 describes the comparative experimental results over 20 datasets and the corresponding analyses. The conclusions and future work are presented in Section 5.

2 Preliminary Knowledge

2.1 Neighborhood Rough Set

The neighborhood rough set was first proposed by Hu et al. [9,30], as an improvement of the conventional rough set. The most significant difference is that the neighborhood rough set is constructed by a neighborhood relation instead of an indiscernibility relation. Therefore, the superiorities of the neighborhood rough set are: (1) it can process data with complicated types; (2) it is equipped with the natural structure of multi-granularity if various radii are used.

Generally, a data can be represented by a triple $DS =< U, AT, d >$, where U is a set of finite samples, AT is a set of condition attributes, and d is a decision [30–32]. d records the labels of samples, $\forall x \in U$ and $\forall a \in AT$, $a(x)$ denotes the value of x over condition attribute a, and $d(x)$ indicates the label of x. According to d, an equivalence relation over U can be obtained as Eq. (1).

$$IND(d) = \{(x, y) \in U \times U:\ d(x) = d(y)\} \tag{1}$$

Following $IND(d)$, a partition $U/IND(d) = \{X_1, X_2, \ldots, X_q\}(q \geq 2)$ over U can be induced. $\forall X_k \in U/IND(d)$, X_k is regarded as the k-th decision class. Particularly, the decision class which contains sample x can also be expressed by $[x]_d$.

Furthermore, suppose that $A \subseteq AT$, given a radius $\delta \geq 0$, a neighborhood relation over U can be derived as:

$$N_A^{\delta} = \{(x, y) \in U \times U:\ dis_A(x, y) \leq \delta\} \tag{2}$$

where dis_A is a distance function between samples x and y with respect to A.

In this study, Euclidean distance is applied, i.e., $dis_A(x,y) = \sqrt{\sum_{a\in A}(a(x)-a(y))^2}$. By N_A^δ, the neighborhood of sample x is then formed such that $\delta_A(x) = \{y \in U:\ dis_A(x,y) \le \delta\}$.

From the viewpoint of Granular Computing (GrC) [10,33,34], both the deriving of $IND(d)$ and N_A^δ are information granulations. The most significant difference between such two information granulations is the inherent mechanism, i.e., the used binary relation. Following these results of information granulations, the definition of neighborhood lower and upper approximations, which are basic units of neighborhood rough set, have also been proposed by Cheng et al. [30].

Definition 1. *Given a data* $DS = \langle U, AT, d\rangle$ *and* δ, $\forall A \subseteq AT$ *and* $\forall X_k \in U/IND(d)$, *the neighborhood lower and upper approximations of* X_k *with respect to* A *are respectively defined as*:

$$\underline{\delta_A}(X_k) = \{x \in U:\ \delta_A(x) \subseteq X_k\} \tag{3}$$

$$\overline{\delta_A}(X_k) = \{x \in U:\ \delta_A(x) \cap X_k \neq \varnothing\} \tag{4}$$

By the above definition, a neighborhood of X_k is obtained, i.e., a pair such that $[\underline{\delta_A}(X_k), \overline{\delta_A}(X_k)]$. In addition, When the lower approximation is not equal to upper approximation, then the pair is called neighborhood rough set.

2.2 Supervised Attribute Reduction

As is well-known, a neighborhood rough set is frequently employed to execute the supervised learning tasks. Furthermore, in many supervised learning tasks, feature selection plays an important role in improving the generalization performance, decreasing the complexity of classifier, and so on. The superiority of attribute reduction is that it can be easily expanded with respect to different requirements in real-world applications. Therefore, various forms of attribute reduction have emerged in recent years. As far as the neighborhood rough set is concerned, the following two measures, approximation quality [30] and conditional entropy [8,19] can be used to further explore the forms of attribute reduction.

Definition 2. *Given a data* $DS = \langle U, AT, d\rangle$ *and* δ, $\forall A \subseteq AT$, *the approximation quality of* d *based on* A *is*:

$$r_A(d) = \frac{\left|\bigcup_{k=1}^{q} \underline{\delta_A}(x_k)\right|}{|U|} \tag{5}$$

where $|\diamond|$ *denotes the cardinality of the set* $\diamond$.

Obviously, $\gamma_A(d) \in [0,1]$ holds. Approximation quality reflects the percentage of the samples which belong to each decision class in lower approximation. Naturally, the higher the value of the approximation quality, the higher the degree of such belongingness.

Definition 3. *Given a data* $DS = \langle U, AT, d\rangle$ *and* δ, $\forall A \subseteq AT$, *the conditional entropy of* d *based on* A *is*:

$$CE_A(d) = -\frac{1}{|U|}\sum\nolimits_{x\in U} \left|\delta_A(x) \cap [x]_d\right| log\frac{\left|\delta_A(x) \cap [x]_d\right|}{|\delta_A(x)|} \tag{6}$$

It is proved that $CE_A(d) \in \left[0, \frac{|U|}{e}\right]$ holds [8]. Conditional entropy reflects the discriminative ability of A relative to d. As the value of the conditional entropy descends, the discriminating ability of A relative to d becomes more and more powerful.

Definition 4. *Given a data* $DS = \langle U, AT, d\rangle$ *and a measure* φ, C_φ *is a constraint related to measure* φ, $\forall A \subseteq AT$, A *is regarded as a* φ*-reduct if and only if*

(1) $\varphi_A(d)$ *meets the constraint* C_φ;
(2) $\forall A' \subset A$, $\varphi_{A'}(d)$ *does not meet the constraint* C_φ.

Without loss of generality, the constraint shown in Def. 4 is closely related to the used measure φ. Suppose that the measure φ is approximation quality shown in Def. 2, then the constraint C_γ is $\gamma_A(d) \geq \gamma_{AT}(d)$; if the form of measure φ is conditional entropy shown in Def. 3, then the constraint C_{CE} is $CE_A(d) \leq CE_{AT}(d)$.

2.3 Unsupervised Attribute Reduction

It is well known that supervised attribute reduction relies heavily on the labels of samples, so it is time-consuming and costly to obtain labels of samples from many real-world tasks. However, unsupervised attribute reduction does not need to obtain such labels, this is why unsupervised attribute reduction has recently been paid much more attention to [14].

For unsupervised attribute reduction, if approximation quality or conditional entropy is still required to serve as a measure, the immediate problem is how to make labels for samples. As it has been pointed out by Qian et al. [35], a pseudo-label strategy can be introduced into the construction of a rough set model. The pseudo labels of samples are generated by using the information over condition attributes. Therefore, it is not difficult to present the following definitions of approximation quality and conditional entropy.

Definition 5. *Given an unsupervised data* $IS = \langle U, AT\rangle$ *and* δ, $\forall A \subseteq AT$, *the unsupervised approximation quality based on* A *is*:

$$\gamma_A = \frac{1}{|A|}\sum\nolimits_{a\in A}\left(\gamma_{A-\{a\}}(d^a)\right) \tag{7}$$

where d^a *is a pseudo-label decision which records the pseudo labels of samples by using condition attribute* a.

A partition can also be obtained such that $U/IND(d^a) = \{X_1^a, X_2^a, \ldots, X_q^a\}$. $\forall X_k^a \in U/IND(d^a)$, X_k^a is the k-th decision class based on the pseudo-label decision d^a.

Similar to the approximation quality shown in Def. 2, $\gamma_A \in [0, 1]$ also holds. Nevertheless, different from Def. 2, the unsupervised approximation quality implies the correlation between a group of attributes and a single attribute. Although the unsupervised approximation quality is not strictly monotonic, it also possesses the property similar to that of Def. 2, i.e., the higher the value of unsupervised approximation quality, the higher the degree of such a correlation.

Definition 6. *Given an unsupervised data* $IS = \langle U, AT\rangle$ *and* δ, $\forall A \subseteq AT$, *the unsupervised conditional entropy based on* A *is*:

$$CE_A = \frac{1}{|A|}\sum\nolimits_{a\in A}\left(CE_{A-\{a\}}(d^a)\right) \tag{8}$$

where d^a *is a derived decision that records the pseudo labels of samples by using condition attribute* a.

According to the definition of conditional entropy shown in Def. 3, $CE_A \in \left[0, \frac{|U|}{e}\right]$ also holds. Note that the unsupervised conditional entropy also indicates the correlation between a set

of attributes and a single attribute. Naturally, it possesses a property similar to that of Def. 3, i.e., the lower the value of unsupervised conditional entropy, the higher the degree of such a correlation.

Definition 7. *Given an unsupervised data* $IS = \langle U, AT \rangle$ *and a measure* φ,C_φ *is a constraint related to measure* φ, $\forall A \subseteq AT$, A *is regarded as a* φ*-reduct if and only if*

(1) φ_A *meets the constraint* C_φ;
(2) $\forall A' \subset AT$, $\varphi_{A'}$ *does not meet the constraint* C_φ.

Without loss of generality, the constraint shown in Def. 7 is closely related to the used measure φ. Suppose that we serve unsupervised approximation quality as the form of measure φ, then the constraint C_γ is $\gamma_A \geq \gamma_{AT}$; if we serve unsupervised conditional entropy as the form of measure φ, then the constraint C_{CE} is $CE_A \leq CE_{AT}$.

3 Proposed Approach

Following Defs. 4 and 7, it is found that attribute reduction is closely related to the used measure. In other words, if different measures are employed then different results of reduct can be generated. From this viewpoint it can be concluded that the used measure is the key to deriving an expected reduct. Nevertheless, most previous measures based on single perspective may fall into the following limitations.

(1) The measure based on single perspective ignores the diverse evaluations [15,16]. The less diverse evaluations, the identified attributes will be less confident. This is mainly because if a measure is fixed to perform attribute reduction, then in the iterations of seeking out reduct, only such a measure is applied to assess the importance of candidate attributes. It follows that such a result of evaluation may be invalidated [36] or ineffective for some other measures.
(2) The measure based on single perspective ignores the complex constraint which is helpful in terminating the procedure of selecting attributes. For instance, given a form of attribute reduction that preserves the value of approximation quality in data, i.e., in Def. 4, the measure φ is $\gamma_{AT}(d)$, such a derived reduct is only equipped with a single characteristic. Immediately, some other measures, such as conditional entropy and unsupervised related measures are not fully taken into account.

From the discussions above, a new measure is proposed to further solve such problems. Since both supervised and unsupervised cases have been presented in the above section, they will be introduced into our new proposed measure.

3.1 Quality-to-Entropy Ratio

Approximation quality and conditional entropy can be used to evaluate the importance of attributes. In Defs. 2 and 6, the relationships between such measures and the importance of attributes are revealed. And these two relationships are at opposite poles. To further unify the relationships, the format of ratio is employed to coordinate two relationships. Moreover, in order to enhance the relationship between conditional entropy and the importance of attributes, an exponential function is added to conditional entropy.

Definition 8. *Given a data* $DS = \langle U, AT, d \rangle$ *and* δ, $\forall A \subseteq AT$, *the quality-to-entropy ratio is*:

$$\kappa_A(d) = \frac{\gamma_A(d)}{\exp(CE_A)} \tag{9}$$

where $\gamma_A(d)$ *is the approximation quality of* d *based on* A *shown in Def. 2,* CE_A *is the unsupervised conditional entropy based on* A *shown in Def. 6.*

From the perspective of attribute reduction, conditional entropy is used to build a bridge between condition attributes and decisions [37]. It frequently reveals a learning relationship between samples and labels in data. In this study, unsupervised data can be labeled through a pseudo-label strategy, and then conditional entropy helps uncover the learning relationships among condition attributes. That is why we use unsupervised conditional entropy instead of unsupervised approximation quality in Eq. (9).

Furthermore, by the form of $\kappa_A(d)$, it is observed that a higher value of approximation quality $\gamma_A(d)$ and a lower value of unsupervised conditional entropy CE_A, then a higher value of the quality-to-entropy ratio. It is also consistent with the previous objectives of attribute reductions related to measures $\gamma_A(d)$ and CE_A. That is, if approximation quality is employed, then a higher value of approximation quality is expected through adding qualified attributes or removing poor attributes; if conditional entropy is employed, then a lower value of approximation quality is expected through adding qualified attributes or removing low-quality attributes.

Theorem 1. *Given a data* $DS = \langle U, AT, d\rangle$ *and* δ, $\forall A \subseteq AT$, $\kappa_A(d) \in [0,1]$.

Proof. By the property of approximation quality, $\gamma_A(d) \in [0,1]$ holds. By the property of conditional entropy, $CE_A \in \left[0, \frac{|U|}{e}\right]$ holds. Immediately, $\exp(CE_A) \in \left[1, e^{\frac{|U|}{e}}\right]$ holds by the form of Eq. (9). In other words, $\kappa_A(d)$ will get the minimal value 0 if $\gamma_A(d) = 0$ and $\exp(CE_A) = e^{\frac{|U|}{e}}$; $\kappa_A(d)$ will get the maximal value 1 if $\gamma_A(d) = 1$ and $\exp(CE_A) = 1$. □

Following Sections 2.1 and 2.2, it is known that the higher the value of the approximation quality, the more influential the discriminating performance of the condition attributes relative to decision d; the lower the value of the unsupervised conditional entropy, the stronger the discriminative power of the condition attributes relative to the pseudo labels. Since CE_A is not strictly monotonically decreasing, we can conclude that $\exp(CE_A)$ is also not strictly monotonically decreasing. To sum up, the higher the value of the quality-to-entropy ratio, the stronger the discriminative power [38,39] of the condition attributes relative to decisions. Furthermore, it should be pointed out that our new measure does not possess the property of strict monotony. An illustrative example is elaborated below.

Example 1. *For example, see Table 1. In Table 1,* $U = \{x_1, x_2, x_3, x_4\}$ *is the universe. Suppose that* $AT = \{a_1, a_2, a_3, a_4\}$ *is the set of condition attributes;* d *is the decision;* $d^{a_1}, d^{a_2}, d^{a_3}$ *and* d^{a_4} *represent the pseudo labels of samples based on four different condition attributes, respectively. They are obtained through using a learning approach. Additionally, the radius is given by 0.2.*

Table 1: An example of data

U	a_1	a_2	a_3	a_4	d	d^{a_1}	d^{a_2}	d^{a_3}	d^{a_4}
x_1	0.2171	0.1477	0.3107	0.2077	1	•	•	•	•
x_2	0.3550	0.4542	0.8767	0.9224	1	◯	•	•	◯
x_3	0.5172	0.8985	0.7375	0.3809	2	◯	◯	•	•
x_4	0.8232	0.8768	0.6847	0.0531	1	•	◯	◯	•

Firstly, the corresponding partitions over U *are*:

$$U/IND\,(d) = \{X_1, X_2\} = \{\{x_1, x_2, x_4\}, \{x_3\}\},$$

$$U/IND\,(d^{a_1}) = \left\{X_1^{a_1}, X_2^{a_1}\right\} = \{\{x_1, x_4\}, \{x_2, x_3\}\},$$

$$U/IND\,(d^{a_2}) = \left\{X_1^{a_2}, X_2^{a_2}\right\} = \{\{x_1, x_2\}, \{x_3, x_4\}\},$$

$$U/IND\,(d^{a_3}) = \left\{X_1^{a_3}, X_2^{a_3}\right\} = \{\{x_1, x_2, x_3\}, \{x_4\}\},$$

$$U/IND\,(d^{a_4}) = \left\{X_1^{a_4}, X_2^{a_4}\right\} = \{\{x_1, x_3, x_4\}, \{x_2\}\}.$$

By Eq. (9), $\kappa_{AT}(d)$ *is then calculated by the following results*:

$$\gamma_{AT}(d) = \frac{\left|\underline{\delta}_{AT}(X_1) \cup \underline{\delta}_{AT}(X_2)\right|}{|U|} = 1,$$

$$CE_{AT} = \frac{1}{|AT|}\left(CE_{AT-\{a_1\}}(d^{a_1}) + CE_{AT-\{a_2\}}(d^{a_2}) + CE_{AT-\{a_3\}}(d^{a_3}) + CE_{AT-\{a_4\}}(d^{a_4})\right)$$

$$= \frac{1}{4}(1 + 1 + 1 + 1) = 1.$$

Therefore, $\kappa_{AT}(d) = \dfrac{\gamma_{AT}(d)}{\exp(CE_{AT})} = 0.3679.$

Presume that the condition attributes a_1 *is removed from data and then* $A = \{a_2, a_3, a_4\}$. *By Eq. (9),* $\kappa_A(d)$ *is calculated by*

$$\gamma_A(d) = \frac{\left|\underline{\delta}_A(X_1) \cup \underline{\delta}_A(X_2)\right|}{|U|} = 1,$$

$$CE_A = \frac{1}{|A|}\left(CE_{A-\{a_2\}}(d^{a_2}) + CE_{A-\{a_3\}}(d^{a_3}) + CE_{A-\{a_4\}}(d^{a_4})\right)$$

$$= \frac{1}{3}(1 + 0.5 + 0.5) = 0.6667.$$

Therefore, $\kappa_A(d) = \dfrac{\gamma_A(d)}{\exp(CE_A)} = \dfrac{1}{e^{0.6667}} = 0.5134.$

Though $A \subseteq AT$ *holds,* $\kappa_{AT}(d) < \kappa_A(d)$, *such a case demonstrates that the new measure* κ *is not strictly monotonic with respect to the number of used condition attributes.*

The quality-to-entropy ratio proposed above presents a form of attribute reduction as follows.

Definition 9. *Given a data* $DS = \langle U, AT, d\rangle$ *and* $\forall A \subseteq AT$, A *is regarded as a* κ*-reduct if and only if*

$$\frac{\kappa_A(d)}{\kappa_{AT}(d)} \geq \theta \tag{10}$$

$$\forall A' \subset A, \quad \frac{\kappa_{A'}(d)}{\kappa_{AT}(d)} < \theta \tag{11}$$

where $\theta \in [0, 1]$ *is a threshold.*

Following Def. 9, it is observed that as a minimal subset of attributes, κ-reduct improves the quality-to-entropy ratio. Immediately, an open problem is how to seek out such a reduct. Without loss of generality, it is frequently required to evaluate the significance or importance of attributes in AT [40,41]. The qualified attributes can be selected into the reduct pool, or low-quality attributes can be removed from the reduct pool. Based on the widely used greedy searching mechanism [42–44], the following definition presents a significant attribute in considering our proposed quality-to-entropy ratio.

Definition 10. *Given a data* $DS = \langle U, AT, d\rangle, \forall A \subseteq AT, \forall a \in AT - A$, *the significance with respect to quality-to-entropy ratio is defined as*:

$$Sig_{\kappa a}(d) = \kappa_{A\cup\{a\}}(d) - \kappa_A(d) \tag{12}$$

The above significance function shows: as the value rises, the condition attribute becomes more and more significant, and such an attribute is highly possible to be added to the reduct pool. For instance, if $Sig_{\kappa a_1}(d) < Sig_{\kappa a_2}(d)$ where $a_1, a_2 \in AT - A$, then $\kappa_{A\cup\{a_1\}}(d) < \kappa_{A\cup\{a_2\}}(d)$. Such a result illustrates that: compared to a_1, if a_2 is selected into A, then the derived quality-to-entropy ratio will be higher.

The above interpretation is consistent with the semantic explanation of attribute reduction shown in Def. 9, i.e., a higher value of the quality-to-entropy ratio is expected.

3.2 Algorithm Description

Based on the significant function shown in Eq. (12), Algorithm 1 is designed to seek out a κ-reduct.

Algorithm 1 Forward Greedy Searching to κ-reduct (FGS-κ)

Input: A data $DS = \langle U, AT, d\rangle$, a radius δ, a threshold θ.
Output: A κ-reduct A.

Let $A = \varnothing$;
Calculate $\gamma_{AT}(d)$ by Eq. (5);
For each $a \in AT$ **do**
    Generate pseudo labels of samples in U by using a;
End
Calculate CE_{AT} by Eq. (8);
Calculate quality-to-entropy ratio $\kappa_{AT}(d)$ by Eq. (9);
Repeat
    For each $a \in AT - A$ **do**
        Calculate $\gamma_{A\cup\{a\}}(d)$;
        Calculate $CE_{A\cup\{a\}}$;
        Calculate $\kappa_{A\cup\{a\}}(d)$;
    End
    Select attribute $b = argmax\{Sig_{\kappa a}(d) : \forall a \in AT - A\}$;
    Let $A = A \cup \{b\}$;
    Calculate $\kappa_A(d)$;
Until $\dfrac{\kappa_A(d)}{\kappa_{AT}(d)} < \theta$;
Return A

To simplify the discussion of the time complexity of Algorithm 1. Firstly, k-means clustering is employed to obtain the pseudo labels of samples. T is the iteration times of k-means, and k is the number of clusters, then the time complexity of producing pseudo labels is $\mathcal{O}(k \cdot T \cdot |U| \cdot |AT|)$. Secondly, $\kappa_{A \bigcup \{a\}}(d)$ is calculated at most $(1 + |AT|) \cdot |AT| / 2$ times. Finally, the time complexity of Algorithm 1 is $\mathcal{O}\left(|U|^2 \cdot |AT|^3 + k \cdot T \cdot |U| \cdot |AT|\right)$.

4 Experimental Analysis

4.1 Data

To demonstrate the performance of our proposed strategy, 18 UCI datasets and 2 Yale face datasets have been employed in this study. Table 2 summarizes the detailed statistics of these data.

Table 2: Data sets descriptions

ID	Data sets	# Samples	# Attributes	# Labels
1	Connectionist bench (Sonar, Mines *vs*. Rocks)	208	60	2
2	Describable textures dataset	5500	40	3
3	Drive-face	606	6400	3
4	First-order theorem proving	6118	51	2
5	Libras movement	360	90	15
6	Lee silverman voice treatment (LSVT) voice rehabilitation	126	256	2
7	Madelon	4400	500	2
8	Molecular biology (Promoter Gene Sequences)	106	57	2
9	OPPORTUNITY activity recognition	2551	242	40
10	Optical recognition of handwritten digits	5620	64	10
11	Ozone level detection	2536	72	15
12	Quality assessment of digital colposcopies	287	63	2
13	Single proton emission computed tomography (SPECT) heart	267	44	2
14	Synthetic control chart time series	600	60	6
15	Ultrasonic flowmeter diagnostics-meter D	180	43	4
16	Urban land cover	675	147	9
17	Waveform database generator (Version 2)	5000	40	3
18	Wisconsin diagnostic breast cancer	569	30	6
19	Yale face (32 × 32)	165	1024	15
20	Yale face (64 × 64)	165	4096	15

4.2 Configuration

All the experiments are carried out on a desktop computer with the Windows 10 operating system, Advanced Micro Devices (AMD) Ryzen 5 5500U with Radeon Graphics (2.10 GHz) and 32.00 GB memory. The programming environment is Matlab R2020a.

In this section, two groups of comparative experiments are designed and executed. For all the experiments, 20 different radii δ have been used, they are $\delta = 0.02, 0.04, \ldots, 0.40$, increasing with the step of 0.02. Furthermore, k-means clustering [44] is adopted in the following experiments to generate the pseudo labels of samples, where the value of k is the same as the number of decision classes in the data.

Moreover, 10-fold cross-validation is used in the experiments, dividing the samples in U into ten groups [45–47]. They are $U_1, U_2, \ldots,$ *and* U_{10}. In the first round of calculation, the set $U_2 \cup \cdots \cup U_{10}$ is regarded as the training set for calculating reduct, *and* U_1 is the set of testing samples; $\cdots$; in the last round of calculation, the set $U_1 \cup \cdots \cup U_9$ is regarded as the training set for calculating reduct, and U_{10} is the set of testing samples. Three classifiers are also employed to test the performances of derived reducts, they are Support Vector Machine (SVM) [48], Classification and Regression Tree (CART) [49] and K-Nearest Neighbor (KNN) [50].

4.3 Experiment (Group 1)

In the first group of the experiment, we will compare our proposed approach with previously supervised attribute reduction and unsupervised attribute reduction strategies. The supervised attribute reduction is based on the approximation quality [30], which is called Supervised Approximation Quality Reduct (SAQR). The unsupervised attribute reduction is based on the conditional entropy, which is called Unsupervised Conditional Entropy Reduct (UCER).

Since 20 different radii have been used to obtain reducts in our experiments, the mean values related to those 20 different reducts are presented in the following subsection.

4.3.1 Comparisons among Classification Stabilities

The classification stabilities [15] derived by different kinds of reducts are compared. Classification stability is used to test the stability of classification results if data perturbation (simulated by cross-validation) happens. Such a computation is based on the distribution of the classification results. The higher the value of classification stability, the better the ability to the data perturbation.

The following Table 3 reports the mean values of classification stabilities obtained over 20 datasets. A closer look at Table 3 reveals the following facts.

(1) Whichever classifier is adopted, the classification stabilities related to κ-reducts are greater than those related to reducts derived by SAQR and UCER in most datasets. Take data "Drive-Face (ID:3)" as an example. Through the use of SVM, the classification stability of FGS-κ, SAQR and UCER are 89.25%, 81.60% and 78.18%, respectively. Through the use of CART, the classification stability of FGS-κ, SAQR and UCER are 70.40%, 64.38% and 61.03%, respectively. Through the use of KNN, the classification stability of FGS-κ, SAQR and UCER are 84.25%, 63.75% and 41.25%, respectively. From this point of view, it is concluded that by comparing with the single measure used in SAQR and UCER, the quality-to-entropy-based multiple measures in FGS-κ helps select attributes with superior data adaptability, i.e., the slight data perturbation in training data will not lead to more significant variation of the classification results.

(2) From the perspective of the average values, the classification stability of FGS-κ is more remarkable than those related to both SAQR and USER. Through the use of SVM, the classification stability of FGS-κ is over 8.24% higher than others. Through the use of CART, the classification stability of FGS-κ is over 8.93% higher than others. Through the use of KNN, the classification stability of FGS-κ is over 13.50% higher than others.

Table 3: The comparisons of classification stabilities (%)

ID	SVM			CART			KNN		
	FGS-κ	SAQR	USER	FGS-κ	SAQR	USER	FGS-κ	SAQR	USER
1	**99.22**	99.01	98.19	**96.66**	96.56	96.14	**94.01**	93.44	92.93
2	**87.73**	77.22	79.32	**79.78**	60.88	66.32	**91.05**	72.22	75.66
3	**89.25**	81.60	78.18	**70.40**	64.38	61.03	**84.25**	63.75	41.25
4	**91.96**	84.56	70.36	**74.64**	72.71	69.42	**89.91**	84.38	72.91
5	**98.80**	91.93	78.98	**87.78**	79.30	72.16	**99.23**	90.85	77.48
6	**99.33**	84.88	83.33	**97.16**	83.03	76.46	**84.00**	75.06	77.47
7	**84.84**	82.26	75.15	**86.58**	83.20	75.00	**66.79**	65.58	60.12
8	**99.86**	89.47	87.83	82.81	**83.43**	78.73	**77.54**	69.78	69.52
9	**84.24**	78.75	76.27	**71.51**	67.87	66.69	**86.06**	78.42	71.15
10	**88.00**	74.68	70.83	**66.25**	64.53	62.00	**82.50**	70.80	62.88
11	**100.00**	100.00	100.00	**94.18**	82.84	81.67	**92.62**	92.51	91.77
12	**97.60**	91.48	85.40	**80.80**	69.08	71.92	**94.40**	75.64	76.44
13	74.34	**83.52**	79.53	**64.96**	53.06	59.88	**77.96**	62.10	57.35
14	**81.90**	59.14	68.19	**73.33**	53.14	70.09	**80.00**	58.85	73.52
15	**100.00**	100.00	100.00	**74.45**	70.62	73.66	**82.58**	78.86	80.66
16	**99.75**	99.68	99.67	**99.62**	99.59	99.45	**99.86**	99.80	99.75
17	72.63	**76.08**	74.66	**85.91**	83.97	73.72	**87.30**	83.44	85.80
18	**91.41**	84.34	82.91	**79.16**	68.06	71.39	**89.65**	75.03	70.76
19	**96.68**	77.63	78.29	**72.50**	59.21	63.40	**87.20**	65.87	64.84
20	**94.54**	69.54	68.36	**59.60**	56.36	59.48	**91.51**	69.63	61.39
Average	**91.60**	84.12	81.94	**79.72**	72.60	72.60	**86.92**	75.17	74.31

From the discussions above, our new approach can select more critical attributes than other attribute reduction strategies, and these selected attributes are able to replace raw attributes to complete subsequent tasks.

4.3.2 Comparisons among Classification Accuracies

The classification accuracies corresponding to different results of reduct are compared. Table 4 reports the average values of classification accuracy obtained over 20 datasets.

Table 4: The comparisons of classification accuracies (%)

ID	SVM			CART			KNN		
	FGS-κ	SAQR	UCER	FGS-κ	SAQR	UCER	FGS-κ	SAQR	UCER
1	**99.12**	97.36	98.74	**98.24**	96.55	98.24	**93.66**	92.10	93.10
2	**82.92**	80.97	70.73	**73.17**	65.34	56.09	**75.61**	75.56	73.17
3	**89.75**	36.25	28.75	**42.50**	27.50	40.25	**80.25**	41.25	48.75
4	**86.56**	83.15	72.63	**76.10**	69.47	73.68	**85.33**	81.05	80.00
5	**97.65**	97.33	93.59	**87.73**	86.21	84.69	**98.49**	98.13	91.63
6	**97.50**	94.16	92.50	**94.83**	90.83	84.16	**88.33**	81.33	88.16
7	**65.37**	62.50	63.88	**77.69**	75.00	79.16	**52.19**	47.22	51.38
8	**80.70**	79.85	80.00	**74.73**	73.68	73.68	**75.43**	71.93	70.17
9	**64.24**	36.36	21.21	**58.48**	46.66	33.33	**41.21**	30.30	12.12
10	**88.00**	66.25	58.75	**56.25**	45.00	36.25	**78.75**	58.75	63.75
11	**82.91**	82.91	82.91	**80.46**	79.08	79.39	**82.31**	79.07	81.28
12	**96.10**	80.00	94.40	**79.20**	64.00	72.00	**88.00**	80.00	88.00
13	**59.81**	54.53	57.69	**71.73**	54.42	70.61	**64.62**	54.03	56.57
14	**70.47**	42.85	61.90	**80.95**	66.66	72.38	**68.57**	38.10	66.67
15	**73.59**	73.58	73.58	**79.25**	73.58	77.35	**69.57**	62.26	62.26
16	**99.91**	99.61	99.82	**99.46**	99.32	99.46	**99.73**	99.64	99.73
17	**61.11**	57.72	58.13	**80.56**	78.08	77.78	**80.56**	69.58	72.22
18	**82.21**	80.00	74.07	**79.26**	77.24	74.82	**78.52**	76.72	64.44
19	**85.86**	47.40	79.30	**73.72**	47.20	69.70	**77.92**	42.80	71.70
20	**74.55**	60.61	36.36	**51.52**	41.70	36.36	**67.88**	54.55	51.52
Average	**81.91**	70.67	69.95	**75.79**	68.28	69.47	**77.35**	66.72	69.28

A closer look at Table 4 reveals the following facts.

(1) Whichever classifier is employed, the accuracies corresponding to κ-reducts are greater than those corresponding to both SQAR and USER. Take the dataset "Optical Recognition of Handwritten Digits (ID:10)" as an example. Through the use of SVM, the classification accuracy of κ-reduct, SAQR and UCER are 88.00%, 66.25% and 58.75%, respectively. Through the use of CART, the classification accuracy of κ-reduct, SAQR and UCER are 56.25%, 45.00% and 36.25%, respectively. Through the use of KNN, the classification accuracy of κ-reduct, SAQR and UCER are 78.75%, 58.75% and 63.75%, respectively. Because of this, it can be found that the attributes selected by our FGS-κ can also provide better learning performance.

(2) Interestingly, compared to others, the κ-reducts derived from FGS-κ perform significantly better over high-dimensional datasets. Take the data "Yale face (64×64) (ID:20)" as an example, which possesses 4096 attributes. The classification accuracy of κ-reduct is 29.04% higher than the reduct derived from SAQR, and 36.58% higher than the reduct derived from USER. For "Connectionist Bench (Sonar, Mines *vs*. Rocks) (ID:1)", which possesses 60 attributes. The classification accuracy of κ-reduct is 1.71% higher than the reduct derived from SAQR, and 0.38% higher than the reduct derived from USER.

(3) From the perspective of the average values, the classification accuracies related to FGS-κ are also more significant than those of both SAQR and USER. Through the use of SVM, the classification accuracy of FGS-κ is over 13.74% higher than others. Through the use of CART, the classification accuracy of FGS-κ is over 8.33% higher than others. Through the use of KNN, the classification accuracy of FGS-κ is over 10.38% higher than others. Such an observation indicates that our proposed algorithm applies to several different classifiers.

4.4 Experiment (Group 2)

In this experiment, with respect to two measures addressed in this study, our proposed approach is compared with five other popular approaches, they are:

(1) The novel rough set algorithm for fast adaptive attribute reduction in classification (FAAR) [51];
(2) The multi-criterion neighborhood attribute reduction (MNAR) [17];
(3) The robust attribute reduction based on rough sets (RARR) [52];
(4) The attribute reduction algorithm based on fuzzy self-information (FSIR) [53];
(5) The attribute reduction based on neighborhood self-information (NSIR) [54].

4.4.1 Comparisons among Classification Stabilities

In this part, the classification stabilities derived from different kinds of reducts are compared. Table 5 shows the mean values of different classification stabilities obtained over 20 datasets. A closer look at Table 5 reveals the following facts.

(1) Whichever classifier is chosen, the classification stabilities based on κ-reducts are more significant than others over most datasets. Take the dataset "Wisconsin Diagnostic Breast Cancer (ID:18)" as an example. Through the use of SVM, the classification stability of FGS-κ, FAAR, MNAR, RARR, FSIR and NSIR are 91.42%, 81.81%, 59.41%, 83.18%, 70.80% and 83.46%, respectively. Through the use of CART, the classification stability of FGS-κ, FAAR, MNAR, RARR, FSIR and NSIR are 79.16%, 75.81%, 76.74%, 72.99%, 61.55% and 71.49%, respectively. Through the use of KNN, the classification stability of FGS-κ, FAAR, MNAR, RARR, FSIR and NSIR are 89.65%, 87.58%, 63.85%, 73.73%, 63.58% and 71.95%, respectively.
(2) From the perspective of the average values, the classification stability of FGS-κ is also greater than those related to FAAR, MNAR, RARR, FSIR and NSIR, respectively. Through the use of SVM, the classification stability of FGS-κ is over 5.87% higher than others. Through the use of CART, the classification stability of FGS-κ is over 8.21% higher than others. Through the use of KNN, the classification stability of FGS-κ is over 8.73% higher than others.

From the discussions above, our new approach is superior to state-of-the-art attribute reduction strategies in offering stable classification results from the discussions above.

Table 5: The comparisons of classification stabilities (%)

Classifier	ID	FGS-κ	FAAR	MNAR	RARR	FSIR	NSIR
SVM	1	99.21	99.12	92.63	**99.26**	99.12	98.83
	2	**87.73**	79.93	80.95	75.14	74.38	85.78

(Continued)

Table 5: Continued

Classifier	ID	FGS-κ	FAAR	MNAR	RARR	FSIR	NSIR
	3	**89.25**	86.45	62.00	85.00	82.15	70.40
	4	**91.95**	89.32	87.01	91.86	90.22	83.78
	5	**98.80**	94.26	69.02	86.58	59.31	85.13
	6	**99.33**	82.20	67.33	65.20	69.06	79.55
	7	**84.84**	83.43	68.61	73.93	71.80	77.37
	8	**99.86**	98.03	97.58	94.00	94.00	94.00
	9	**84.24**	83.67	70.91	78.30	81.75	77.87
	10	**88.00**	81.60	59.50	86.55	81.80	70.30
	11	**100.00**	100.00	100.00	100.00	100.00	100.00
	12	**97.60**	87.48	90.40	96.04	91.28	94.52
	13	74.35	76.54	72.20	**83.66**	69.46	71.07
	14	**81.91**	59.38	69.52	61.52	75.81	74.57
	15	**100.00**	100.00	100.00	100.00	100.00	100.00
	16	99.75	97.92	97.93	**100.00**	99.27	99.50
	17	**72.64**	68.73	69.77	71.67	68.98	67.41
	18	**91.42**	81.81	59.41	83.18	70.80	83.46
	19	**96.68**	92.38	71.34	87.46	76.35	89.32
	20	**94.55**	83.79	63.03	73.76	83.61	66.94
	Average	**91.61**	86.30	77.46	84.66	81.95	83.49
CART	1	**96.66**	91.58	88.42	91.79	91.70	91.93
	2	**79.78**	63.69	65.71	59.21	63.00	70.47
	3	**61.25**	56.80	59.25	61.09	57.70	54.40
	4	**74.46**	72.56	74.38	60.68	60.38	71.66
	5	**87.79**	87.33	66.60	83.72	57.51	76.72
	6	**97.17**	76.07	63.33	63.21	70.40	76.50
	7	**86.58**	66.18	65.55	65.83	70.05	65.70
	8	**82.81**	80.43	74.82	76.94	73.91	73.91
	9	**67.88**	65.48	59.69	67.30	62.66	62.84
	10	**66.25**	57.80	55.00	63.70	60.60	65.00
	11	**94.18**	92.29	88.40	93.80	82.54	83.05
	12	**80.80**	72.48	74.00	75.60	76.28	78.52
	13	**64.96**	59.71	58.87	58.12	58.16	59.04
	14	**73.33**	57.38	67.62	67.81	69.33	69.61
	15	**74.45**	71.18	72.59	74.02	71.92	71.74
	16	**99.62**	99.56	99.56	98.65	99.26	99.53
	17	**85.92**	84.44	77.77	81.11	85.12	84.69
	18	**79.16**	75.81	76.74	72.99	61.55	71.49
	19	**72.50**	71.10	60.62	68.44	66.76	71.27
	20	**59.61**	54.12	55.45	58.78	58.06	58.03
	Average	**79.26**	72.80	70.22	72.14	69.85	72.81
KNN	1	94.01	87.31	**97.40**	87.58	91.70	86.63
	2	**91.05**	63.69	65.71	59.21	63.00	70.47

(Continued)

Table 5: Continued

Classifier	ID	FGS-κ	FAAR	MNAR	RARR	FSIR	NSIR
	3	**84.25**	79.90	62.25	72.40	71.80	64.15
	4	**89.91**	88.54	85.04	88.21	83.85	85.73
	5	**99.23**	89.33	70.87	79.91	58.70	82.16
	6	**84.00**	78.15	66.66	70.30	70.88	76.08
	7	**66.79**	65.39	64.16	61.68	62.04	58.26
	8	**77.54**	76.17	74.48	72.06	70.93	71.05
	9	**86.06**	83.63	76.36	74.90	76.66	74.97
	10	**82.50**	74.60	60.75	72.65	68.85	63.05
	11	**92.62**	90.35	76.97	82.46	91.86	89.04
	12	**94.40**	78.32	81.60	81.08	79.56	84.64
	13	**77.96**	75.19	71.69	81.14	71.42	76.45
	14	**80.00**	59.81	69.52	65.90	75.42	75.14
	15	**82.58**	74.00	73.38	76.11	75.63	74.18
	16	**99.86**	89.89	89.98	94.09	99.57	99.59
	17	**87.31**	84.13	83.33	86.25	84.98	86.75
	18	**89.65**	87.58	63.85	73.73	63.58	71.95
	19	**87.20**	77.22	58.34	69.86	68.64	73.80
	20	**91.52**	84.39	68.48	72.87	69.51	68.03
	Average	**86.92**	79.38	73.04	76.12	74.93	76.61

4.4.2 Comparisons among Classification Accuracies

In this part, the classification accuracies based on different kinds of reducts are also compared. Different classification accuracies acquired by 20 datasets are shown in Table 6. A closer look at Table 6 reveals the following facts.

(1) Whichever classifier is chosen, the classification accuracies based on κ-reducts are more significant than others over most datasets. Take the dataset "Ultrasonic Flowmeter Diagnostics-Meter D (ID:15)" as an example. By employing SVM, the classification accuracy of FGS-κ, FAAR, MNAR, RARR, FSIR and NSIR are 73.58%, 71.48%, 70.58%, 71.48%, 71.48% and 71.48%, respectively. By employing CART, the classification accuracy of FGS-κ, FAAR, MNAR, RARR, FSIR and NSIR are 79.24%, 79.20%, 75.18%, 72.22%, 77.78% and 72.22%, respectively. By employing KNN, the classification accuracy of FGS-κ, FAAR, MNAR, RARR, FSIR and NSIR are 69.56%, 59.56%, 55.18%, 62.22%, 57.78% and 62.22%, respectively.

(2) From the perspective of the average values, the classification precision related to FGS-κ is also more significant than those of FAAR, MNAR, RARR, FSIR and NSIR. By employing SVM, the classification accuracy of FGS-κ is over 5.92% higher than others. By employing CART, the classification accuracy of FGS-κ is over 8.46% higher than others. By employing KNN, the classification accuracy of FGS-κ is over 5.61% higher than others.

From the discussions above, our new approach is superior to popular attribute reduction strategies in providing more precise classification results.

Table 6: The comparisons of classification accuracies (%)

Classifier	ID	FGS-κ	FAAR	MNAR	RARR	FSIR	NSIR
SVM	1	**99.12**	99.12	95.61	99.12	99.12	99.12
	2	**82.92**	71.42	78.57	78.57	54.76	18.09
	3	**89.75**	85.00	10.00	16.25	1.25	21.25
	4	**86.57**	74.21	78.43	82.56	78.43	78.43
	5	**97.65**	95.02	24.73	13.61	9.07	78.73
	6	**97.50**	95.00	75.00	74.17	16.67	90.00
	7	**65.36**	63.67	50.00	56.94	13.89	63.88
	8	80.70	81.03	80.05	**81.32**	79.20	79.31
	9	**64.24**	61.67	58.21	59.39	12.12	58.48
	10	**88.00**	83.75	62.50	57.50	63.75	61.25
	11	**82.91**	80.27	79.77	81.50	82.75	87.74
	12	**96.01**	72.00	72.00	92.00	94.00	94.00
	13	**59.80**	56.20	54.97	53.59	58.87	55.73
	14	**70.76**	57.13	38.95	47.62	38.09	70.43
	15	**73.58**	71.48	70.58	71.48	71.48	71.48
	16	**99.90**	99.50	99.46	87.64	97.36	98.91
	17	**61.11**	58.22	53.89	57.79	49.42	56.89
	18	**82.21**	80.96	71.48	79.25	62.96	81.22
	19	**85.86**	84.90	52.90	50.60	61.13	85.40
	20	**74.54**	68.78	42.44	59.39	68.18	68.49
	Average	**81.91**	77.09	62.69	64.96	55.71	70.94
CART	1	**98.24**	92.10	90.35	92.98	91.28	91.22
	2	**73.17**	66.19	70.95	71.81	59.52	66.19
	3	**42.50**	35.00	40.97	25.00	27.50	31.25
	4	**76.10**	72.56	72.56	66.77	73.38	70.08
	5	**87.74**	85.89	76.96	75.62	69.07	78.73
	6	**94.83**	88.33	76.67	69.66	73.33	81.66
	7	**77.69**	62.50	47.22	59.72	71.11	66.67
	8	74.74	63.79	**75.86**	71.03	70.69	70.69
	9	**48.45**	48.37	42.42	44.39	45.15	47.51
	10	**46.25**	42.50	35.00	30.00	21.25	25.00
	11	**80.46**	78.08	75.85	75.47	78.33	79.70
	12	**79.20**	68.00	74.00	72.00	78.00	72.00
	13	**71.73**	64.50	66.19	68.95	68.87	70.73
	14	**80.95**	47.61	76.19	52.38	38.09	71.43
	15	**79.24**	79.20	75.18	72.22	77.78	72.22
	16	**99.45**	99.49	98.36	93.91	99.27	99.18
	17	**80.56**	73.33	76.11	76.11	72.23	77.78

(Continued)

Table 6: Continued

Classifier	ID	FGS-κ	FAAR	MNAR	RARR	FSIR	NSIR
	18	**79.26**	76.29	60.00	74.07	57.78	79.26
	19	**73.72**	72.70	46.80	47.00	33.90	73.60
	20	**51.57**	44.54	44.54	48.85	42.21	45.45
	Average	**74.79**	68.05	66.24	64.15	60.95	68.51
KNN	1	**93.67**	91.61	90.47	91.14	87.64	92.86
	2	**75.61**	70.95	73.19	73.19	59.52	70.95
	3	**80.25**	73.75	47.50	47.50	26.25	73.75
	4	**85.37**	83.38	83.38	81.73	78.08	78.08
	5	**98.49**	97.76	75.71	76.06	69.07	86.56
	6	**88.33**	85.00	69.16	72.50	67.50	83.00
	7	**52.19**	48.22	45.27	49.89	42.50	50.22
	8	**75.44**	65.52	74.14	72.75	72.41	65.52
	9	**41.21**	40.42	39.39	39.94	37.71	41.21
	10	**78.75**	73.75	55.00	25.00	31.25	73.75
	11	**82.31**	73.67	80.46	80.54	80.46	73.67
	12	**88.00**	84.00	80.00	84.00	88.00	84.00
	13	**64.61**	53.32	56.20	61.55	58.87	57.32
	14	**68.57**	66.19	38.09	42.85	38.09	68.43
	15	**69.56**	59.56	55.18	62.22	57.78	62.22
	16	**99.73**	99.55	99.55	87.18	99.64	99.55
	17	**80.56**	76.11	71.67	78.89	72.22	76.11
	18	**78.52**	77.04	48.89	73.33	68.15	77.04
	19	**77.92**	74.60	44.80	47.60	35.80	73.60
	20	**67.88**	66.67	54.54	45.46	18.18	66.67
	Average	**77.35**	73.06	64.19	64.66	59.83	72.73

5 Conclusions and Future Work

In such a study, with a review of limitations derived from measures based on single perspective, a novel measure, called quality-to-entropy, is proposed. The superiorities of our research are: in the process of searching reduct, not only can the relationship between condition attributes and labels be characterized from the supervised perspective, but also the internal structure of data can be quantitatively described. More importantly, our new measure can be further introduced into other search strategies. Through extensive experiments, it has been demonstrated that the proposed measure helps select attributes with significantly superior classification performance.

Further investigations will be focused on the following aspects:

(1) Our strategy is developed based on the neighborhood rough set. It can also be further introduced into other rough sets, e.g., fuzzy rough set [36,53,55] and decision-theoretic rough set [41].
(2) Fused measures may increase time consumption for seeking out a reduct, accelerators [3,42] to reduce the corresponding elapsed time are then necessary.

Funding Statement: This work was supported by the National Natural Science Foundation of China (Grant Nos. 62006099, 62076111), the Key Research and Development Program of Zhenjiang-Social Development (Grant No. SH2018005), the Natural Science Foundation of Jiangsu Higher Education (Grant No. 17KJB520007), Industry-school Cooperative Education Program of the Ministry of Education (Grant No. 202101363034).

Conflicts of Interest: The authors declare that they have no conflicts of interest in reporting regarding the present study.

References

[1] Z. H. Jiang, X. B. Yang, H. L. Yu, D. Liu, P. X. Wang *et al.,* "Accelerator for multi-granularity attribute reduction," *Knowledge-Based Systems*, vol. 177, pp. 145–158, 2019.

[2] Y. Chen, X. B. Yang, J. H. Li, P. X. Wang and Y. H. Qian, "Fusing attribute reduction accelerators," *Information Sciences*, vol. 587, pp. 354–370, 2022.

[3] P. X. Wang and X. B. Yang, "Three-way clustering method based on stability theory," *IEEE Access*, vol. 9, pp. 33944–33953, 2021.

[4] Q. Chen, T. H. Xu and J. J. Chen, "Attribute reduction based on lift and random sampling," *Symmetry-Basel*, vol. 14, no. 9, pp. 1828, 2022.

[5] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[6] X. B. Yang, Y. Qi, D. J. Yu, H. L. Yu and J. Y. Yang, "α-Dominance relation and rough sets in interval-valued information systems," *Information Sciences*, vol. 294, pp. 334–347, 2015.

[7] K. Y. Liu, X. B. Yang, H. L. Yu, H. Fujita, X. J. Chen *et al.,* "Supervised information granulation strategy for attribute reduction," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 9, pp. 2149–2163, 2020.

[8] X. Zhang, C. L. Mei, D. G. Chen and J. H. Li, "Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy," *Pattern Recognition*, vol. 56, pp. 1–15, 2016.

[9] Q. H. Hu, W. Pedrycz, D. R. Yu and J. Lang, "Selecting discrete and continuous features based on neighborhood decision error minimization," *IEEE Transactions on Systems, Man, and Cybernetics, Part B(Cybernetics)*, vol. 40, no. 1, pp. 137–150, 2010.

[10] Y. Chen, K. Y. Liu, J. J. Song, H. Fujita, X. B. Yang *et al.,* "Attribute group for attribute reduction," *Information Sciences*, vol. 535, no. 5, pp. 64–80, 2020.

[11] K. Y. Liu, X. B. Yang, H. Fujita, D. Liu, X. Yang *et al.,* "An efficient selector for multi-granularity attribute reduction," *Information Sciences*, vol. 505, pp. 457–472, 2019.

[12] Z. H. Jiang, K. Y. Liu, X. B. Yang, H. L. Yu, H. Fujita *et al.,* "Accelerator for supervised neighborhood based attribute reduction," *International Journal of Approximate Reasoning*, vol. 119, pp. 122–150, 2020.

[13] Z. Yuan, H. M. Chen, T. R. Li, Z. Yu, B. B. Sang *et al.,* "Unsupervised attribute reduction for mixed data based on fuzzy rough sets," *Information Sciences*, vol. 572, no. 1, pp. 67–87, 2021.

[14] X. B. Yang and Y. Y. Yao, "Ensemble selector for attribute reduction," *Applied Soft Computing*, vol. 70, pp. 1–11, 2018.

[15] Y. N. Chen, P. X. Wang, X. B. Yang and H. L. Yu, "BEE: Towards a robust attribute reduction," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 12, pp. 3927–3962, 2022.

[16] J. Z. Li, X. B. Yang, X. N. Song, J. H. Li, P. X. Wang *et al.,* "Neighborhood attribute reduction: A multi-criterion approach," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 4, pp. 731–742, 2019.

[17] J. H. Dai, W. T. Wang, H. W. Tian and L. Liu, "Attribute selection based on a new conditional entropy for incomplete decision systems," *Knowledge-Based Systems*, vol. 39, no. 1, pp. 207–213, 2013.

[18] J. H. Dai, Q. Xu, W. T. Wang and H. W. Tian, "Conditional entropy for incomplete decision systems and its application in data mining," *International Journal of General Systems*, vol. 41, no. 74, pp. 713–728, 2012.

[19] Y. H. Qian, J. Y. Liang, W. Pedrycz and C. Y. Dang, "Positive approximation: An accelerator for attribute reduction in rough set theory," *Artificial Intelligence*, vol. 174, no. 9–10, pp. 597–618, 2010.
[20] W. Y. Li, H. M. Chen, T. R. Li, J. H. Wan and B. B. Sang, "Unsupervised feature selection via self-paced learning and low-redundant regularization," *Knowledge-Based Systems*, vol. 240, no. 11, pp. 108150, 2022.
[21] X. B. Yang, Y. Qi, H. L. Yu, X. N. Song and J. Y. Yang, "Updating multi-granulation rough approximations with increasing of granular structures," *Knowledge-Based Systems*, vol. 64, pp. 59–69, 2014.
[22] S. P. Xu, X. B. Yang, H. L. Yu, D. J. Yu and J. Y. Yang, "Multi-label learning with label-specific feature reduction," *Knowledge-Based Systems*, vol. 104, pp. 52–61, 2016.
[23] K. Y. Liu, T. R. Li, X. B. Yang, H. R. Ju, X. Yang *et al.,* "Hierarchical neighborhood entropy based multi-granularity attribute reduction with application to gene prioritization," *International Journal of Approximate Reasoning*, vol. 148, no. 3, pp. 57–67, 2022.
[24] J. Qian, D. Q. Miao, Z. H. Zhang and W. Li, "Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation," *International Journal of Approximate Reasoning*, vol. 52, no. 2, pp. 212–230, 2011.
[25] K. Y. Liu, X. B. Yang, H. L. Yu, J. S. Mi, P. X. Wang *et al.,* "Rough set based semi-supervised feature selection via ensemble selector," *Knowledge-Based Systems*, vol. 165, pp. 282–296, 2019.
[26] X. B. Yang, M. Zhang, H. L. Dou and J. Y. Yang, "Neighborhood systems-based rough sets in incomplete information system," *Knowledge-Based Systems*, vol. 24, no. 6, pp. 858–867, 2011.
[27] Y. H. Qian, H. H. Cheng, J. T. Wang, J. Y. Liang, W. Pedrycz *et al.,* "Grouping granular structures in human granulation intelligence," *Information Sciences*, vol. 382–383, no. 6, pp. 150–169, 2017.
[28] X. B. Yang, Y. H. Qian and J. Y. Yang, "On characterizing hierarchies of granulation structures via distances," *Fundamenta Informaticae*, vol. 123, no. 3, pp. 365–380, 2013.
[29] Q. H. Hu, D. R. Yu and Z. X. Xie, "Neighborhood classifiers," *Expert Systems with Applications*, vol. 34, no. 2, pp. 866–876, 2008.
[30] K. Cheng, S. Gao, W. L. Dong, X. B. Yang, Q. Wang *et al.,* "Boosting label weighted extreme learning machine for classifying multi-label imbalanced data," *Neurocomputing*, vol. 403, no. 8, pp. 360–370, 2020.
[31] S. P. Xu, H. R. Ju, L. Shang, W. Pedrycz, X. B. Yang *et al.,* "Label distribution learning: A local collaborative mechanism," *International Journal of Approximate Reasoning*, vol. 121, no. 7, pp. 59–84, 2020.
[32] Y. Chen, P. X. Wang, X. B. Yang, J. S. Mi and D. Liu, "Granular ball guided selector for attribute reduction," *Knowledge-Based Systems*, vol. 229, pp. 107326, 2021.
[33] X. B. Yang, S. P. Xu, H. L. Dou, X. N. Song, H. L. Yu *et al.,* "Multi-granulation rough set: A multiset based strategy," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 277–292, 2017.
[34] X. B. Yang, S. C. Liang, H. L. Yu, S. Gao and Y. H. Qian, "Pseudo-label neighborhood rough set: Measures and attribute reductions," *International Journal of Approximate Reasoning*, vol. 105, pp. 112–129, 2019.
[35] Y. H. Qian, J. Y. Liang and W. Wei, "Consistency-preserving attribute reduction in fuzzy rough set framework," *International Journal of Machine Learning and Cybernetics*, vol. 4, no. 4, pp. 287–299, 2013.
[36] T. H. Xu, G. Y. Wang and J. Yang, "Finding strongly connected components of simple digraphs based on granulation strategy," *International Journal of Approximate Reasoning*, vol. 118, no. 1, pp. 64–78, 2020.
[37] Y. Y. Yao and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Information Sciences*, vol. 179, no. 7, pp. 867–882, 2009.
[38] C. Z. Wang, Q. H. Hu, X. Z. Wang, D. G. Chen, Y. H. Qian *et al.,* "Feature selection based on neighborhood discrimination index," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 2986–2999, 2018.
[39] X. S. Rao, K. Y. Liu, J. J. Song, X. B. Yang and Y. H. Qian, "Gaussian kernel fuzzy rough based attribute reduction: An acceleration approach," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 1, pp. 1–17, 2020.
[40] J. J. Song, E. C. C. Tsang, D. G. Chen and X. B. Yang, "Minimal decision cost reduct in fuzzy decision-theoretic rough set model," *Knowledge-Based Systems*, vol. 126, no. 1, pp. 104–112, 2017.
[41] Y. H. Qian, J. Y. Liang, W. Pedrycz and C. Y. Dang, "An efficient accelerator for attribute reduction from incomplete data in rough set framework," *Pattern Recognition*, vol. 44, no. 8, pp. 1658–1670, 2011.

[42] X. S. Rao, X. B. Yang, X. Yang, X. J. Chen, D. Liu *et al.,* "Quickly calculating reduct: An attribute relationship-based approach," *Knowledge-based Systems*, vol. 200, pp. 106014, 2020.
[43] Y. X. Liu, Z. C. Gong, K. Y. Liu, S. P. Xu, H. R. Ju *et al.,* "A q-learning approach to attribute reduction," *Applied Intelligence*, vol. 53, pp. 3750–3765, 2022.
[44] P. X. Wang, H. Shi, X. B. Yang and J. S. Mi, "Three-way k-means: Integrating k-means and three-way decision," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2767–2777, 2019.
[45] J. Ba, K. Y. Liu, H. R. Ju, S. P. Xu, T. H. Xu *et al.,* "Triple-g: A new mgrs and attribute reduction," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 2, pp. 337–356, 2022.
[46] Z. C. Gong, Y. X. Liu, T. H. Xu, P. X. Wang and X. B. Yang, "Unsupervised attribute reduction: Improving effectiveness and efficiency," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 11, pp. 3645–3662, 2022.
[47] C. -C. Chang and C. -J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27, 2011.
[48] M. Krzywinski and N. Altman, "Classification and regression trees," *Nature Methods*, vol. 14, no. 8, pp. 755–756, 2017.
[49] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. 24, no. 7, pp. 750–753, 1975.
[50] S. Y. Xia, H. Zhang, W. H. Li, G. Y. Wang, E. Giem *et al.,* "GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp. 1231–1242, 2020.
[51] L. J. Dong, D. G. Chen, N. L. Wang and Z. H. Lu, "Key energy-consumption feature selection of thermal power systems based on robust attribute reduction with rough sets," *Information Sciences*, vol. 532, no. 5, pp. 61–71, 2020.
[52] C. Z. Wang, Y. Huang, W. P. Ding and Z. H. Cao, "Attribute reduction with fuzzy rough self-information measures," *Information Sciences*, vol. 549, no. 12, pp. 68–86, 2021.
[53] C. Z. Wang, Y. Huang, M. W. Shao, Q. H. Hu and D. Chen, "Feature selection based on neighborhood self-information," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 4031–4042, 2020.
[54] J. J. Song, X. B. Yang, X. N. Song, H. L. Yu and J. Y. Yang, "Hierarchies on fuzzy information granulations: A knowledge distance based lattice approach," *Journal of Intelligent & Fuzzy Systems*, vol. 27, no. 3, pp. 1107–1117, 2014.
[55] W. W. Yan, J. Ba, T. H. Xu, H. L. Yu, J. L. Shi *et al.,* "Beam-influenced attribute selector for producing stable reduct," *Mathematics*, vol. 10, no. 4, pp. 553, 2022.