IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 December 2023, accepted 30 December 2023, date of publication 2 January 2024,
date of current version 8 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3349270

RESEARCH ARTICLE

Effective Attribute Reduction Algorithm Based on Fuzzy Uncertainties Using Shared Neighborhood Granulation

SHENGLI GAO
Faculty of Intelligent Engineering Technology, Jiangsu Vocational College of Finance and Economics, Huaian 223003, China
e-mail: shengligao@jscj.edu.cn

ABSTRACT As a very prominent research application of the theory of rough sets, attribute reduction technique has made significant strides in a lot of fields, including decision making, granular computing, etc. In particular, fuzzy attribute reduction approaches contribute greatly in the presence of uncertain data. However, most of fuzzy relations used in these approaches lack the discriminant ability to sample similarity, failing to identify the feature significance satisfactorily. In this article, a novel scheme using the shared neighborhood fuzzy uncertainties is proposed. Firstly, the concept of shared neighborhood is formulated, and then employed to establish the fuzzy similarity relation that effectively captures the sample similarity. Secondly, two fuzzy uncertainty measures named joint entropy and discrimination index based on shared neighborhood fuzzy relation are defined, which can quantify the feature's significance to the uncertainty characterization. Finally, two heuristic searching algorithms are designed to identify reducts aimed at minimizing the fuzzy uncertainties. Some comparative studies are investigated to examine the advantage of the designed reduction algorithms in classifier modeling. The reported analyses on public data sets verify that the designed algorithms outperform some representative and latest algorithms.

INDEX TERMS Attribute reduction, fuzzy neighborhood rough set, granular computing, rough set theory, uncertainty measure.

I. INTRODUCTION

In 1982, Professor Z. Pawlak coined the rough set theory which is generally acknowledged as an efficient and relatively new mathematical implement to process the incomplete, inaccurate, and undefinable data [1], [2], [3]. In rough set theory, no extra information is needed, thus it has gained huge attention from considerable number of researchers across a lot of fields, to name a few, artificial intelligence, decision making, machine learning, granular computing, data mining [4], [5], [6].

Toady, the rapid developments of information technologies have resulted the exponential growth of data attributes. That lies some core challenges ahead machine learning modeling. Consequently, over the past thirty years, more and more theories, methodologies and techniques of rough set are devoted to the exploration and exploitation of attribute reduction [7], [8], [9]. As a crucial application of rough set theory, attribute reduction is also termed as variable selection or feature reduction, and it is a common but significant pre-processing way for big data analysis [10]. Its prime target is to define and find a minimal subset of condition attributes termed as reduct with higher dependency and lower redundancy [11]. As we know, an ideal reduct enables to maintain physical semantics, boost learning accuracy, accelerate computational speed, provide model interpretability, etc [12], [13], [14]. Up to now, various approaches have been developed for different marching learning problems. For instance, considering the technique of attribute reduction in multi-label problem, Liu et al. [15] designed a neighborhood granulation attribute reduction algorithm that fuses various concepts; Chen et al. [16] used label specific features algorithm and

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Hao Chen.

sample section strategy with randomness to devise a new type of algorithm of feature reduction. Considering the ensemble learning problem, Wang et al. [17] introduced the forest optimization algorithm into the process of picking up reduct which can return multiple reducts, and used these reducts to develop an ensemble framework for executing voting classification over testing samples. Considering the monotonic classification problem, Zhang et al. [18] applied the matrix approach for lower approximation in an inconsistent decision system to give the discriminative concept tree with the relations by dominance, and then fused the evaluation functions by tree approach to establish an efficient algorithm of searching lower approximation reduct. Considering the semi-supervised learning problem, Liu et al. [19] proposed a semi-supervised attribute reduction approach that can handle the partially labeled data with label propagation algorithm and ensemble selector.

Recently, due to the capability of analysing fuzzy, real-valued, and even mix-valued data, fuzzy rough attribute reduction approaches become popular by the assistance of fuzzy rough techniques. The existing methods mainly focus on how to establish the fuzzy rough models, e.g., some operators for computing fuzziness and fitness functions for estimating features. However, less attention has been paid to the adaptative generation of fuzzy similarity relation. This oversight can lead to inadequacies in the selected reducts for constructing learning models. To address this issue, a novel scheme based on shared neighborhood fuzzy uncertainties are presented. Firstly, we introduce the concept of shared neighborhood, using it to obtain the fuzzy similarity relation. The shared neighborhood aims to extract the samples located within the neighborhoods of two given samples simultaneously, which allows us to capture the sample similarity precisely. Secondly, as the shared neighborhood fuzzy relation is defined, fuzzy granulation technique with stronger discriminating ability can be induced, and then two fuzzy uncertainty measures named joint entropy and discrimination index based on shared neighborhood fuzzy relation are defined, which can quantify the feature's significance to the uncertainty characterization. Finally, two heuristic searching algorithms are designed to find the reducts which are required to minimize the fuzzy uncertainties, respectively.

In what follows, the contributions of this article are clarified.

1) A novel fuzzy granulation technique using shared neighborhood fuzzy relation is proposed, which leverages the concept of shared neighborhood to distinguish the sample similarity.
2) Two uncertainty measures of joint entropy and discrimination index on the basis of shared neighborhood fuzzy relation are defined. These measures are applied to exploit the inherent relationship between condition-to-decision attribute. Notably, such two measures are different from the previous expressions, their utilization of shared neighborhood fuzzy relation instead of crisp neighborhood relation enhances the capacity to identify feature quality.
3) An attribute reduction framework aimed at minimizing the aforementioned uncertainty measures is devised. And the reported analyses suggest the superiority of our designed attribute reduction algorithms over several competing approaches.

II. RELATED WORK

Attribute reduction is a vibrant research problem that attracts numerous techniques to solve it. Specially, fuzzy and fuzzy-rough approaches provide powerful solutions to this problem from the perspective of uncertainty [20], [21]. This section mainly reviews methods closely related to fuzzy and fuzzy-rough attribute reduction.

To the best of our knowledge, there are two categories of fuzzy and fuzzy-rough attribute reduction approaches.

1) One way is to create new fuzzy measurement as feature evaluation metric. For example, Hu et al. [22] adjusted the fuzzy rough approximation with probability, and then proposed a theory about fuzzy probabilistic approximation spaces that develop fuzzy information measures for attribute reduction. Rao et al. [23] presented a very quick scheme of feature reduction through taking multiple Gaussian kernels into account in which multiple levels of granularity by different scales of fuzzy granules were considered. Liu et al. [24] used fuzzy technique for fuzzy relevance and redundancy to appraise the importance of semi-supervised features in partial decision system also called partially labeled data, and developed a semi-supervised feature selection. Deng et al. [25] realized a feature reduction method based on dual-granulation in both feature and label space for label distribution decision system by fuzzy rough information measurements. Guo et al. [26] proposed a new attribute evaluation function named double fuzzy consistency that characterizes the uncertainties together in both fuzzy upper and lower approximations for attribute reduction.

2) The other way is to follow the original expression of approximations while promoting the fuzzy and fuzzy-rough models. For instance, Jensen et al. [27] revisited the metric function of similarity, and devised several fuzzy rough set-based ways of feature evaluation. Wang et al. [28] redefined the space of fuzzy rough approximations by irreflexive and symmetric structure, and presented a concept of approximation quality degree for characterizing the decision error rate and gave a new evaluation fitness function to assess the goodness of the observed features. An et al. [29] designed a new metric called relative distance and used it to define the relative fuzzy similarity relation in the lower and upper approximations for attribute reduction. Such an approach can successfully fit a complex data distribution where class densities have huge differences. Tan et al. [30] created a new approximation model driven by roughness and fuzziness, and induced significance measures to evaluate the identification ability for attribute reduction.

An et al. [31] developed a robust fuzzy rough set based on kNN granules and then followed the approximation expression to design a fuzzy-rough-uncertainty-based semi-supervised feature selection. More information can be found in [32] that recalled several important foundations related to the model of fuzzy rough theory, and then investigated so many advanced corresponding attribute reduction schemes comprehensively.

III. BASIC KNOWLEDGE

Subsequently, we will recall some useful and necessary notations in fuzzy rough set theory, some of which are meaningful for introducing our approach.

Let S and A be a set of samples (always called the universe) and a set of attributes which are often used to describe the samples, respectively. The pair of (S, A) is called an information table or system. Specially, if $A = C \cup D$ and $C \cap D = \emptyset$ where C is a set of condition attribute, and D is a decision attribute, then this special type of information table or system is more often termed as a decision table or system written as $T = (S, C \cup D)$ [33].

Let T be a supervised and complete table. Herein, let $R_{C'}$ be a fuzzy binary relation obtained via a subset of condition attributes such that $C' \subseteq C$. Usually, we write it as a $n \times n$ matrix denoted by

$$M(R_{C'}) = \begin{bmatrix} R_{C'}(s_1, s_1) & R_{C'}(s_1, s_2) & \dots & R_{C'}(s_1, s_n) \\ R_{C'}(s_2, s_1) & R_{C'}(s_2, s_2) & \dots & R_{C'}(s_2, s_n) \\ \dots & \dots & \dots & \dots \\ R_{C'}(s_n, s_1) & R_{C'}(s_n, s_2) & \dots & R_{C'}(s_n, s_n) \end{bmatrix} \quad (1)$$

where $R_{C'}(s_i, s_j)$ indicates the relation value between each two samples such as $s_i, s_j \in S$. In this article, we discuss the fuzzy similarity relation. That means, $R_{C'}$ meets the conditions of symmetry and reflexivity. Formally, 1) reflexivity: $\forall s_i \in S$, $R_{C'}(s_i, s_i) = 1$; 2) symmetry: $\forall s_i, s_j \in S$, $R_{C'}(s_i, s_j) = R_{C'}(s_j, s_i)$ [34].

Such a fuzzy relation shown in Eqn. (1) allows $S/R_{C'}$ which are called a collection of fuzzy granules [34]. $\forall s_i \in S$, its fuzzy granule is mathematically represented as

$$[s_i]_{C'} = \frac{R_{C'}(s_i, s_1)}{s_1} + \frac{R_{C'}(s_i, s_2)}{s_2} + \ldots + \frac{R_{C'}(s_i, s_n)}{s_n}. \quad (2)$$

Similar to the generation of fuzzy granule, using information by D, a fuzzy collection of sample s_i is expressed as

$$[s_i]_D = \frac{D(s_i, s_1)}{s_1} + \frac{D(s_i, s_2)}{s_2} + \ldots + \frac{D(s_i, s_n)}{s_n}, \quad (3)$$

where $D(s_i, s_j) = 1$ if s_i is assigned with the same decision attribute to s_i; otherwise, $D(s_i, s_j)$ equals to 0.

Specially, we can calculate the cardinality of Eqns. (2) and (3) such that

$$|[s_i]_{C'}| = \sum_{s_j \in S} R_{C'}(s_i, s_j). \quad (4)$$

$$|[s_i]_D| = \sum_{s_j \in S} D(s_i, s_j). \quad (5)$$

IV. OUR METHOD

How to generate an appropriate fuzzy relation to describe the sample similarity is always an open line, inviting researchers to develop a number of methods. Towards this end, we develop a novel approach in this section.

Definition 1: Considering that $d_{C'}(s_i, s_j)$ is a distance metric, as the Euclidean function is always used, it denotes the Euclidean distance between samples $s_i, s_j \in S$ related to $C' \subseteq C$ in this work, and δ be a radius. $\forall s_i \in S$, the δ-based neighborhood of s_i can be calculated by

$$N^{\delta}_{C'}(s_i) = \{s_j \in S : d_{C'}(s_i, s_j) \leq \delta\}. \quad (6)$$

Based on the concept of radius-based neighborhood, we present the concept of shared neighborhood. In the following, The detailed formulation and definition is exhibited.

Definition 2: Let T be a supervised and complete table. Assume that radius δ is assigned, $\forall s_i, s_j \in S$, the shared neighborhood of s_i and s_j according to $C' \subseteq C$ is defined as

$$SN^{\delta}_{C'}(s_i, s_j) = N^{\delta}_{C'}(s_i) \cap N^{\delta}_{C'}(s_j). \quad (7)$$

Theorem 1: Let T be a supervised and complete table, and $B_1 \subseteq B_2 \subseteq C$. And then we can have $SN^{\delta}_{B_2}(s_i, s_j) \subseteq SN^{\delta}_{B_1}(s_i, s_j)$.

Proof of Theorem 1: $\forall s_i, s_j \in S$, it is simple to get $d_{B_1}(s_i, s_j) \leq d_{B_2}(s_i, s_j)$ if $B_1 \subseteq B_2 \subseteq C$. Correspondingly, with a fixed radius δ, $N^{\delta}_{B_2}(s_i) \subseteq N^{\delta}_{B_1}(s_i)$. Similarly, $N^{\delta}_{B_2}(s_j) \subseteq N^{\delta}_{B_1}(s_j)$. So, we can know that $SN^{\delta}_{B_2}(s_i, s_j) \subseteq SN^{\delta}_{B_1}(s_i, s_j)$.

Theorem 2: Let T be a supervised and complete table. Assume that $\delta_1 \leq \delta_2$ are two different radii assigned to the samples $s_i, s_j \in S$, then $SN^{\delta_1}_{C'}(s_i, s_j) \subseteq SN^{\delta_2}_{C'}(s_i, s_j)$.

Proof of Theorem 2: $\forall s_i, s_j \in S$, with a fixed attribute subset $C' \subseteq C$, we have $N^{\delta_1}_{C'}(s_i) \subseteq N^{\delta_2}_{C'}(s_i)$. Similarly, $N^{\delta_1}_{C'}(s_j) \subseteq N^{\delta_2}_{C'}(s_j)$. It follows that $SN^{\delta_1}_{C'}(s_i, s_j) \subseteq SN^{\delta_2}_{C'}(s_i, s_j)$.

Compared with the traditional neighborhood expression in Eqn. (6), the shared neighborhood may be better for describing the proximity. Not only that, the fuzziness in neighborhood should be exploited. For such purpose, a new fuzzy relation $R^{\delta}_{C'}$ using shared neighborhood can be constructed, in which the fuzzy relation value is specifically computed.

Definition 3: Let T be a supervised and complete table. Assume that radius δ is assigned, the shared neighborhood fuzzy relation value related to $C' \subseteq C$ can be computed by

$$R^{\delta}_{C'}(s_i, s_j) = \begin{cases} 1 & s_i = s_j \\ 0 & SN^{\delta}_{C'}(s_i, s_j) = \emptyset \\ \dfrac{\sum_{s_k \in SN^{\delta}_{C'}(s_i, s_j)} w(s_k, s_i) + w(s_k, s_j)}{2 \cdot |SN^{\delta}_{C'}(s_i, s_j)|} & \text{otherwise} \end{cases} \quad (8)$$

where $w(s_i, s_j)$ indicates the similarity of s_i and s_j formulated by

$$w(s_i, s_j) = \frac{1}{1 + d_{C'}(s_i, s_j)}. \tag{9}$$

Theorem 3: The proposed shared neighborhood fuzzy relation $R^{\delta}_{C'}$ satisfies reflexivity and symmetry.

Proof of Theorem 3: Theorem 3 is direct to prove.

Obviously, $R^{\delta}_{C'}$ is a fuzzy similarity relation. Furthermore, we use this shared neighborhood fuzzy relation to explore the underlying relationship of condition-to-decision attribute. Herein, the uncertainty between fuzzy granule and fuzzy decision class are focused on. Specifically, two measures, i.e., shared neighborhood fuzzy joint entropy and discrimination index are developed to quantify the uncertainty. The detailed definitions and formulas are given.

Definition 4: Let T be a supervised and complete table. The joint entropy in terms of $C' \subseteq C$ is defined as

$$JE(C') = -\frac{1}{|S|} \sum_{s_i \in S} \log \frac{|[s_i]_{R^{\delta}_{C'}} \cap [s_i]_D|}{|S|}. \tag{10}$$

where

$$|[s_i]_{R^{\delta}_{C'}} \cap [s_i]_D| = \sum_{s_j \in S} \min(R^{\delta}_{C'}(s_i, s_j), D(s_i, s_j)). \tag{11}$$

Theorem 4: Let T be a supervised and complete table. Assume that $B_1 \subseteq B_2 \subseteq C$, then we can have $JE(B_1) \leq JE(B_2)$.

Proof of Theorem 4: According to Proof 1, we know that $SN^{\delta}_{B_2}(s_i, s_j) \subseteq SN^{\delta}_{B_1}(s_i, s_j)$, it is easily known that $R^{\delta}_{B_2}(s_i, s_j) \leq R^{\delta}_{B_1}(s_i, s_j)$. Therefore, as $[s_i]_D$ does not change, $[s_i]_{R^{\delta}_{B_2}} \cap [s_i]_D| \leq [s_i]_{R^{\delta}_{B_1}} \cap [s_i]_D|$. It follows that $JE(B_1) \leq JE(B_2)$.

Definition 5: Let T be a supervised and complete table. The discrimination index in terms of $C' \subseteq C$ can be expressed by

$$I(C') = \log \frac{|S|^2}{|R^{\delta}_{C'}|}, \tag{12}$$

where

$$|R^{\delta}_{C'}| = \sum_{s_i \in S} |[s_i]_{R^{\delta}_{C'}}|. \tag{13}$$

Theorem 5: Let T be a supervised and complete table. Assume that $B_1 \subseteq B_2 \subseteq C$, then we can have $I(B_1) \leq I(B_2)$.

Proof of Theorem 5: As proven in Proof 1, $SN^{\delta}_{B_2}(s_i, s_j) \subseteq SN^{\delta}_{B_1}(s_i, s_j)$, it is easily known that $|R^{\delta}_{B_2}| \leq |R^{\delta}_{B_1}|$. It follows that $I(B_1) \leq I(B_2)$.

Obviously, as the discrimination index is monotonous according to the changes of feature subset, $I(C') \in [0, \log |S|]$. Specially, $\max I(C') = \log |S|$ if $|R^{\delta}_{C'}| = |S|$ and $\min I(C') = 0$ if $|R^{\delta}_{C'}| = |S|^2$.

Definition 6: Let T be a supervised and complete table. The discrimination index in terms of $C' \subseteq C$ related to D can be expressed by

$$DI(C') = \log \frac{|R^{\delta}_{C'}|}{|R^{\delta}_{C'} \cap R_D|}, \tag{14}$$

where

$$|R^{\delta}_{C'} \cap R_D| = \sum_{s_i \in S} |[s_i]_{R^{\delta}_{C'}} \cap [s_i]_D|. \tag{15}$$

Usually, a satisfactory condition attribute subset is supposed to granulate samples with high similarity into the same class. Such a subset can guarantee a consistency between fuzzy granule described by condition attributes and fuzzy class described by decision attribute. Ideally, samples located in a fuzzy granule are assigned with the same decision attribute. That means, we expect that the condition attribute subset would make $|[s_i]_{R^{\delta}_{C'}} \cap [s_i]_D|$ in Eqn. (10) and $|R^{\delta}_{C'} \cap R_D|$ in Eqn. (12) as high as possible. It should stress that, unlike the original expressions of entropy and discrimination index, the redefined measures in Eqns. (10) and (12) use the proposed shared neighborhood fuzzy relation instead of traditional neighborhood relation. Therefore, these new measures allow fuzzy operation for the evaluation of feature quality, and enable better discrimination ability in uncertainty.

To enhance the comprehension of these proposed measures, the detailed calculation processes are outlined in the following.

Let us compute the joint entropy and discrimination index on the whole condition attribute set in a synthetic supervised and complete system as exhibited in Table 1.

TABLE 1. A synthetic supervised and complete decision system.

	c_1	c_2	c_3	c_4	c_5	d
s_1	0.1564	0.5061	0.1843	0.2004	0.0388	1
s_2	0.0842	0.6201	0.6209	0.0972	0.0177	1
s_3	0.1664	0.6362	0.2431	0.2349	0.0678	1
s_4	0.0582	0.1685	0.1373	0.0370	0.0023	2
s_5	0.0545	0.1609	0.0954	0.0360	0.0033	2

As shown in Tab. 1, we need to compute on five samples described by five condition attributes, i.e., $\{c_1, c_2, c_3, c_4, c_5\}$. In the following, neighborhood radius δ is appointed as 0.2. Firstly, $N^{0.2}_C(s_1) = \{s_1, s_3\}$, $N^{0.2}_C(s_2) = \{s_2\}$, $N^{0.2}_C(s_3) = \{s_1, s_3\}$, $N^{0.2}_C(s_4) = \{s_4, s_5\}$, $N^{0.2}_C(s_5) = \{s_4, s_5\}$. Then, we can obtain the shared neighborhoods, such as $SN^{0.2}_C(s_1, s_1) = \{s_1, s_3\}$, $SN^{0.2}_C(s_1, s_2) = \emptyset$, $SN^{0.2}_C(s_1, s_3) = \{s_1, s_3\}$, $SN^{0.2}_C(s_1, s_4) = \emptyset$, $SN^{0.2}_C(s_1, s_5) = \emptyset$. According to Eqn. (8), we can induce the similarities, such as $R^{0.2}_C(s_1, s_1) = 1$, $R^{0.2}_C(s_1, s_2) = 0$, $R^{0.2}_C(s_1, s_3) = (1 + 0.8695 + 0.8695 + 1)/4 = 0.9347$, $R^{0.2}_C(s_1, s_4) = 0$, $R^{0.2}_C(s_1, s_5) = 0$. Finally, it is easy to compute the joint entropy and discrimination index: $JE(C) = 1.5470$, $DI(C) = \log \frac{1+0.9370+1+0.9370+1+1+0.9795+0.9795+1}{1+0.9370+1+0.9370+1+1+0.9795+0.9795+1} = 0$.

Correspondingly, it is plain to see that, if the candidate subset of condition attributes could provide us with the lower

the value of the uncertainty measure, then we would regard it as a better one with more discriminate ability.

Moreover, we use these two evaluation measurements for attribute reduction. The general flowchart is provided in the following.

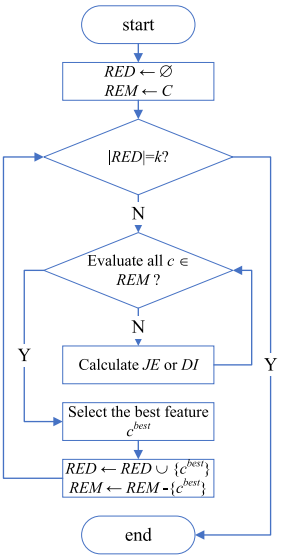

FIGURE 1. Flowchart of the proposed algorithms.

Correspondingly, the complete algorithm, i.e., pseudo codes clarifying the detailed searching process is given.

Algorithm 1 SNFJE (and SNFDI).

Input: A supervised and complete decision system T, radius δ, a predefined selected number k.

Output: A qualified reduct RED.

1: $RED \leftarrow \emptyset$, $REM \leftarrow C$
2: **while** $|RED| \neq k$ **do**
3: **for all** $c \in REM$ **do**
4: **case 1:** Calculate $JE(RED \cup \{c\})$
5: **case 2:** Calculate $DI(RED \cup \{c\})$
6: **end for**
7: **case 1:** Select the best feature c^{best} that provides the minimal value of $JE(RED \cup \{c\})$
8: **case 2:** Select the best feature c^{best} that provides the minimal value of $DI(RED \cup \{c\})$
9: $RED \leftarrow RED \cup \{c^{best}\}$, $REM \leftarrow RED - \{c^{best}\}$
10: **end while**
11: **return** RED.

Our proposed method consists of two stages: the evaluation of candidate features and the selection of qualified features. The first stage (steps 3-6) is to query the significance or quality of each feature, while the second stage (steps 7-9) is to compare them, and determine the most important one. In each iteration of these steps, qualified feature is added into selected pool one by one, and these steps are iterated until the termination condition is satisfied, i.e., k condition attributes are selected. Notably, if the criterion of joint entropy is used, then the case 1 of Algorithm 1 is chosen, and we call the designed algorithm SNFJE; if the criterion of discrimination index is used, then the case 2 of Algorithm 1 is executed, and the corresponding algorithm is named SNFDI. Furthermore, our algorithm adopts a forward greedy searching mode, leading to a computational complexity of Algorithm 1 similar to that of other sequential selection approaches, i.e., $O(|S|^2 \times |C| \times k)$.

TABLE 2. Characteristics of the used data sets.

ID	Data sets	$\lvert S\rvert$	$\lvert C\rvert$
1	Diabetic Retinopathy Debrecen	1151	19
2	Forest Type Mapping	523	27
3	Hill-Valley	1212	100
4	Ionosphere	351	34
5	Leaf	430	15
6	Page Blocks Classification	5473	10
7	Seeds	210	7
8	Speaker Accent Recognition	329	12
9	Wine Quality_Red	1599	11
10	Wine Quality_White	4898	11
11	Wisconsin Diagnostic Breast Cancer	569	30
12	Yeast	1484	8

V. EXPERIMENTS

Naturally, it is necessary to make the simulation experiments to test the proposed SNFJE and SNFDI. For this purpose, we do the experiments on MATLAB. In what follows, the received results are reported and analyzed.

A. CONFIGURATIONS

Twelve public data sets are employed from UCI. Some useful and necessary descriptions are exhibited.

In the following experiments, we mainly compare SNFJE and SNFDI with

1) ALL: It is the naive approach using all the original features of data sets.
2) HANDI [34]: It is a heuristic reduct searching approach using the neighborhood discrimination index as feature evaluation function. It is more efficient in time cost as compared with traditional neighborhood-based attribute reduction since it only computes the cardinality of a neighborhood relation instead of multiple neighborhood similarity classes. The used neighborhood relation is still crisp which may fail to reflect the actual certainty in data.
3) NSIFS [35]: It is a forward greedy approach that applies the neighborhood self-information and entropy measures. It extends some information measures to neighborhood-based ones and can describe the feature significance from an uncertain perspective. Nonetheless, its time efficiency is poor because it requires to calculate the neighborhood classes of all samples.
4) PLAQR [36]: It is a pseudo-labelling decision-theoretic rough set-based reduction algorithm. It mainly uses K-means clustering to produce the pseudo labels which is helpful for supporting the generation of

indistinguishable relation. However, K-means-based pseudo-labelling is not always stable, which may deteriorate the reduction process.

5) SGNRS [37]: It is a neighborhood rough set-based strategy using the supervised granulation. Its neighborhood relation or granule can be more accuracy as supervised neighborhood granulation is able to filter out more negative samples. But such an approach may be easily influenced by the weak supervision like noisy labels in real-world applications.
6) N3Y [38]: It is a neighborhood uncertainty-based feature selection procedure that applies relevancy, redundancy, and granularity interactivity to evaluate feature jointly. It can better alleviate the bias resulted by a fixed neighborhood radius. However, it evaluates features individually, which ignores the overall performance of the selected subset.

For a fair comparison, radius is set as 0.15 in these approaches. To avoid random influences, we utilize the strategy of 10-fold cross-validation. That means, we average these values, and report them in the following.

B. COMPARISONS ON CLASSIFICATION

We adopt the ranking strategy to critique the goodness or quality of selected features by observing CART, KNN and SVM classification performance. Specifically, we feed 10%, 20%, 30%, …, 100% ratios of top-ranking features selected by algorithms to these three classifiers, and the corresponding classification metrics are averaged to report.

In Figure 2 to 4, the comparisons of these algorithms on obtained results classification accuracies are presented. In detail, three types of classification accuracies are focused on.

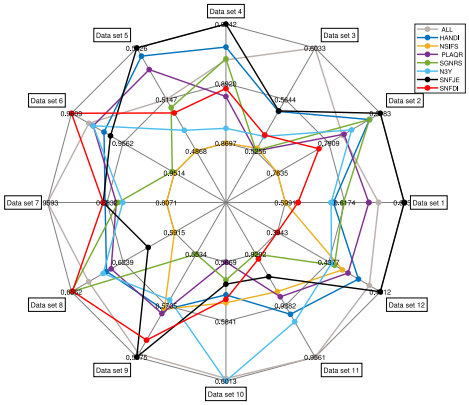

FIGURE 2. Comparison on CART accuracies.

As exhibited in Figure 2 to 4, algorithm with better performance intends to be closer to the boundary of radar chart. So, we can see that SNFJE or SNFDI always achieves the best classification performance. Specifically, in Figure 2, for CART classification accuracies, SNFJE and SNFDI attain

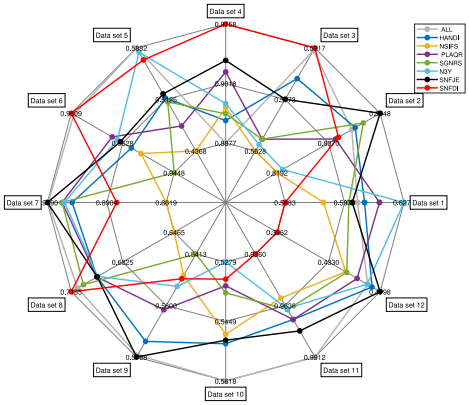

FIGURE 3. Comparison on KNN accuracies.

the maximal accuracies in 8/12 data sets (i.e., Data set 1, 2, 4, 5, 6, 8, 9, 12); for KNN classification accuracies, SNFJE and SNFDI attain the maximal accuracies in 8/12 data sets (i.e., Data set 1, 2, 4, 6, 7, 8, 9, 12); for SVM classification accuracies, SNFJE and SNFDI attain the maximal accuracies in 8/12 data sets (i.e., Data set 2, 3, 4, 7, 8, 9, 12). It should be noticed that, although our proposed methods are sometimes defeated by ALL (i.e., all the original features), it is still evidently superiority to other attribute reduction algorithms.

Specially, although HANDI and SGNRS use the discrimination index and joint entropy, SNFJE and SNFDI are more effective. Such result demonstrates that shared neighborhood fuzzy relation is better than neighborhood relation used in HANDI and SGNRS, mainly because it is more suitable for describing the similarity and fuzziness in neighborhood.

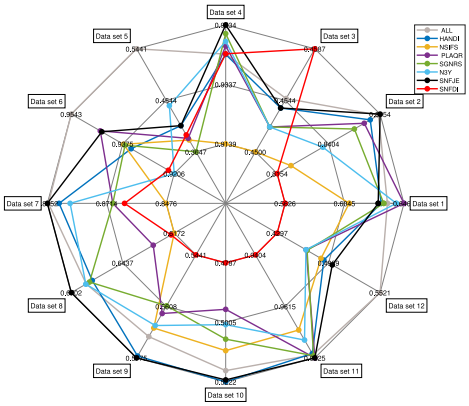

FIGURE 4. Comparison on SVM accuracies.

In Tables 3 to 5, the obtained results of CART, KNN and SVM recalls are presented. It should be pointed out that due to the multi-class data sets, the metric of macro-recall is employed. Similar to the results of classification accuracies, SNFJE or SNFDI always achieves the best classification

TABLE 3. Results of CART macro-recalls.

ID	ALL	HANDI	NSIFS	PLAQR	SGNRS	N3Y	SNFJE	SNFDI
1	0.6276	0.6146	0.5995	0.6247	0.6169	0.6072	**0.6357**	0.6025
2	0.7624	0.7892	0.7548	0.7799	0.8005	0. 7913	**0.8008**	0.7753
3	0.6088	0.5628	0.5267	0.5240	0.5243	**0.6100**	0.5572	0.5399
4	0.9003	0.9060	0.8661	0.8802	0.9028	0.9038	**0.9128**	0.8809
5	0.4964	0.4990	0.4539	0.5095	0.4850	0.4985	**0.5158**	0.4695
6	**0.7843**	0.7094	0.6599	0.7000	0.6495	0.6417	0.6935	0.7654
7	0.8675	**0.8807**	0.7923	0.8599	0.8580	0.8723	0.8765	0.8705
8	0.5418	0.6104	0.4599	0.5867	**0.6743**	0.6641	0.5516	0.6257
9	0.3506	0.3474	0.3444	0.3420	0.3249	0.3380	0.3507	**0.3621**
10	0.3709	0.3495	0.3540	0.3428	0.3472	0.3411	0.3472	**0.3734**
11	**0.9467**	0.9360	0.9291	0.9300	0.9153	0.9161	0.9242	0.9172
12	0.3820	0.3826	0.3387	0.3692	0.3616	0.2527	**0.3923**	0.2619

TABLE 4. Results of KNN macro-recalls.

ID	ALL	HANDI	NSIFS	PLAQR	SGNRS	N3Y	SNFJE	SNFDI
1	**0.6057**	0.6090	0.5878	0.6154	0.6010	0.5999	0.6027	0.5677
2	**0.8673**	0.8380	0.8062	0.8285	0.8452	0.8316	0.8485	0.8085
3	0.5825	0.5917	0.5701	0.5745	0.5745	0.5839	0.5838	**0.5963**
4	0.8406	0.8431	0.8495	0.8551	0.8441	0.8573	0.8641	**0.8780**
5	0.5244	0.4865	0.4114	0.4625	0.4863	0.5302	0.5051	**0.5309**
6	0.6825	0.5868	0.6090	0.6403	0.5962	0.5985	0.5998	**0.7251**
7	0.9007	0.8926	0.8410	0.8906	0.8906	0.8849	**0.9041**	0.8704
8	**0.7539**	0.6241	0.5134	0.5628	0.6811	0.5713	0.6051	0.6301
9	0.3222	0.3328	0.3068	0.3202	0.3150	0.3144	**0.3343**	0.3200
10	0.3251	0.3114	0.3060	0.2969	0.2971	0.3141	0.3127	**0.3258**
11	**0.9844**	0.9645	0.9503	0.9640	0.9597	0.9789	0.9682	0.9247
12	0.3802	0.3791	0.3465	0.3787	0.3681	0.3517	**0.3857**	0.2959

TABLE 5. Results of SVM macro-recalls.

ID	ALL	HANDI	NSIFS	PLAQR	SGNRS	N3Y	SNFJE	SNFDI
1	0.6369	0.6334	0.6100	0.6304	0.6337	0.6217	**0.6481**	0.5580
2	0.8596	0.8539	0.8034	0.8470	0.8513	0.8366	**0.8601**	0.7497
3	0.5050	0.5060	0.5023	0.5045	0.5045	0.5070	0.5063	**0.5113**
4	0.9348	0.9228	0.8863	0.9290	0.9353	0.9318	**0.9433**	0.9257
5	**0.5574**	0.4620	0.4051	0.4244	0.3738	0.4377	0.4427	0.3893
6	**0.5631**	0.4020	0.4317	0.4811	0.4432	0.4566	0.4845	0.3338
7	0.8822	0.8823	0.8446	0.8603	0.8603	0.8635	**0.8874**	0.8419
8	0.4300	0.4255	0.3410	0.3680	0.4328	0.4246	**0.4401**	0.3272
9	0.2843	0.2819	0.2807	0.2787	0.2806	0.2701	**0.2865**	0.2722
10	0.2163	0.2120	0.2078	0.1897	0.1996	0.2005	**0.2178**	0.1752
11	**0.9826**	0.9795	0.9635	0.9848	0.9789	0.9763	0.9778	0.9244
12	**0.4344**	0.3714	0.3334	0.3451	0.3413	0.3548	0.3673	0.2770

performance with respect to macro-recalls. Specifically, for CART macro-recalls, SNFJE and SNFDI attain the maximal accuracies in 7/12 cases; for KNN macro-recalls, SNFJE and SNFDI attain the maximal accuracies in 8/12 cases; for SVM macro-recalls, SNFJE and SNFDI attain the maximal accuracies in 8/12 cases. All the compared algorithms depend on the fundamental concept of neighborhood, while the description techniques are different.

On the whole, SNFJE generally produces more satisfactory classification performance than SNFDI. Hence, we hypothesize that concerning the problem of feature evaluation and selection, the concept of shared neighborhood fuzzy may be more suitable for the formulation of joint entropy instead of discrimination index. Such a phenomenon impels us to explore a more applicable fuzzy relation to improve the discrimination index.

In summary, it is easily known that SNFJE and SNFDI beat the other four algorithms as they yield the outperforming classification results in most data sets. As they are able to identify informative features which are more helpful for constructing classification models, it may be a promising tool in several practical applications of high-dimensional data analysis, including: 1) gene selection for microarray data; 2) band selection for cancer classification by pixel classification in hyperspectral imagery data.

TABLE 6. Comparisons on time complexity.

Algorithm	Time complexity
SNFDI	$O(\|S\|^2 \times \|C\| \times k)$
SNFJE	$O(\|S\|^2 \times \|C\| \times k)$
N3Y	$O(\|S\|^2 \times \|C\| \times k)$
SGNRS	$O(\|S\|^2 \times \|C\| \times k)$
PLAQR	$O(\|S\|^2 \times \|C\| \times k)$
NSIFS	$O(\|S\| \times k)$
HANDI	$O(\|S\|^2 \times (\|C\| \times k + k)/2)$

C. COMPARISONS ON TIME COMPLEXITY

As shown in Table 6, the time complexities of these compared algorithms are listed. Evidently, their time complexities are similar, because all of these compared algorithms employ the same searching framework, i.e., sequentially forward greedy which is the most popular and efficient basic searching strategy. In addition, accelerated methods are expected for a fast version of our proposed methods.

VI. CONCLUSION

This paper introduces a novel fuzzy similarity relation for attribute reduction. This relation allows the concept of shared neighborhood to characterize the sample similarity, thereby enabling the realization of an effective fuzzy granulation technique. Then, the shared neighborhood fuzzy relation is applied to engage with the fuzzy decision class to explore the uncertainty of candidate condition attribute subset. Furthermore, two uncertainty measures called joint entropy and discrimination index are defined. Aiming at minimizing such two measures, two forward sequential searching based attribute reduction algorithms are devised.

The reported experimental results underscore that the proposed SNFJE and SNFDI are superior to various representative and latest attribute reduction algorithms. However, specific limitations remain to be addressed:

- Only simulation experiment is performed to demonstrate the advantage of the proposed SNFJE and SNFDI, but theoretical analysis is lacked.
- Computing shared neighborhood maybe time-consuming, some accelerated methods such as parallel or distributed computing like various equipment and instruments can be applied.
- The proposed algorithms are verified on conventional structured and real-valued data, but it cannot process heterogeneous, multi-source, multi-view and even multi-label data. It is desirable to develop and extend them to a generalized framework.

REFERENCES

[1] A. Skowron and D. Slezak, "Rough sets turn 40: From information systems to intelligent systems," in *Proc. 17th Conf. Comput. Sci. Intell. Syst. (FedCSIS)*, Sep. 2022, pp. 23–34.
[2] H. Qi, Y. Shi, X. Mu, and M. Hou, "Knowledge granularity for continuous parameters," *IEEE Access*, vol. 9, pp. 89432–89438, 2021.
[3] T. Xie and Z. Gong, "A hesitant soft fuzzy rough set and its applications," *IEEE Access*, vol. 7, pp. 167766–167783, 2019.
[4] K. Liu, T. Li, X. Yang, X. Yang, D. Liu, P. Zhang, and J. Wang, "Granular cabin: An efficient solution to neighborhood learning in big data," *Inf. Sci.*, vol. 583, pp. 189–201, Jan. 2022.
[5] K. Liu, T. Li, X. Yang, X. Yang, and D. Liu, "Neighborhood rough set based ensemble feature selection with cross-class sample granulation," *Appl. Soft Comput.*, vol. 131, Dec. 2022, Art. no. 109747.
[6] N. Nandhini and K. Thangadurai, "An incremental rough set approach for faster attribute reduction," *Int. J. Inf. Technol.*, vol. 15, no. 2, pp. 1–15, Feb. 2023.
[7] J. Zhang, K. Liu, X. Yang, H. Ju, and S. Xu, "Multi-label learning with relief-based label-specific feature selection," *Int. J. Speech Technol.*, vol. 53, no. 15, pp. 18517–18530, Aug. 2023.
[8] L. Sun, L. Wang, W. Ding, Y. Qian, and J. Xu, "Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 19–33, Jan. 2021.
[9] L. Sun, T. Yin, W. Ding, Y. Qian, and J. Xu, "Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 5, pp. 1197–1211, May 2022.
[10] L. Sun, T. Yin, W. Ding, and J. Xu, "Hybrid multilabel feature selection using BPSO and neighborhood rough sets for multilabel neighborhood decision systems," *IEEE Access*, vol. 7, pp. 175793–175815, 2019.
[11] T. Yin, H. Chen, Z. Yuan, T. Li, and K. Liu, "Noise-resistant multilabel fuzzy neighborhood rough sets for feature subset selection," *Inf. Sci.*, vol. 621, pp. 200–226, Apr. 2023.
[12] Y. Chen, X. Yang, J. Li, P. Wang, and Y. Qian, "Fusing attribute reduction accelerators," *Inf. Sci.*, vol. 587, pp. 354–370, Mar. 2022.
[13] S. Yang, H. Zhang, G. Shi, and Y. Zhang, "Attribute reductions of quantitative dominance-based neighborhood rough sets with A-stochastic transitivity of fuzzy preference relations," *Appl. Soft Comput.*, vol. 134, Feb. 2023, Art. no. 109994.
[14] M. Akram, H. S. Nawaz, and M. Deveci, "Attribute reduction and information granulation in Pythagorean fuzzy formal contexts," *Expert Syst. Appl.*, vol. 222, Jul. 2023, Art. no. 119794.
[15] J. Liu, Y. Lin, W. Ding, H. Zhang, C. Wang, and J. Du, "Multi-label feature selection based on label distribution and neighborhood rough set," *Neurocomputing*, vol. 524, pp. 142–157, Mar. 2023.
[16] Q. Chen, T. Xu, and J. Chen, "Attribute reduction based on lift and random sampling," *Symmetry*, vol. 14, no. 9, p. 1828, Sep. 2022.
[17] J. Wang, Y. Liu, J. Chen, and X. Yang, "An ensemble framework to forest optimization based reduct searching," *Symmetry*, vol. 14, no. 6, p. 1277, Jun. 2022.
[18] J. Zhang, X. Zhang, and W. Xu, "Lower approximation reduction based on discernibility information tree in inconsistent ordered decision information systems," *Symmetry*, vol. 10, no. 12, p. 696, Dec. 2018.
[19] K. Liu, X. Yang, H. Yu, J. Mi, P. Wang, and X. Chen, "Rough set based semi-supervised feature selection via ensemble selector," *Knowl.-Based Syst.*, vol. 165, pp. 282–296, Feb. 2019.
[20] S. An, Q. Xiao, C. Wang, and S. Zhao, "Granularity self-information based uncertainty measure for feature selection and robust classification," *Fuzzy Sets Syst.*, vol. 470, Oct. 2023, Art. no. 108658.
[21] S. An, X. Guo, C. Wang, G. Guo, and J. Dai, "A soft neighborhood rough set model and its applications," *Inf. Sci.*, vol. 624, pp. 185–199, May 2023.
[22] Q. Hu, D. Yu, Z. Xie, and J. Liu, "Fuzzy probabilistic approximation spaces and their information measures," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 191–201, Apr. 2006.
[23] X. Rao, K. Liu, J. Song, X. Yang, and Y. Qian, "Gaussian kernel fuzzy rough based attribute reduction: An acceleration approach," *J. Intell. Fuzzy Syst.*, vol. 39, no. 1, pp. 679–695, Jul. 2020.
[24] K. Liu, T. Li, X. Yang, H. Chen, J. Wang, and Z. Deng, "SemiFREE: Semisupervised feature selection with fuzzy relevance and redundancy," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 10, pp. 3384–3396, Oct. 2023.
[25] Z. Deng, T. Li, D. Deng, K. Liu, P. Zhang, S. Zhang, and Z. Luo, "Feature selection for label distribution learning using dual-similarity based neighborhood fuzzy entropy," *Inf. Sci.*, vol. 615, pp. 385–404, Nov. 2022.
[26] Y. Guo, M. Hu, X. Wang, E. C. C. Tsang, D. Chen, and W. Xu, "A robust approach to attribute reduction based on double fuzzy consistency measure," *Knowl.-Based Syst.*, vol. 253, Oct. 2022, Art. no. 109585.
[27] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Aug. 2009.
[28] C. Wang, Y. Qian, W. Ding, and X. Fan, "Feature selection with fuzzy-rough minimum classification error criterion," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 2930–2942, Aug. 2022.
[29] S. An, E. Zhao, C. Wang, G. Guo, S. Zhao, and P. Li, "Relative fuzzy rough approximations for feature selection and classification," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2200–2210, Apr. 2023.
[30] A. Tan, W.-Z. Wu, Y. Qian, J. Liang, J. Chen, and J. Li, "Intuitionistic fuzzy rough set-based granular structures and attribute subset selection," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 3, pp. 527–539, Mar. 2019.
[31] S. An, M. Zhang, C. Wang, and W. Ding, "Robust fuzzy rough approximations with kNN granules for semi-supervised feature selection," *Fuzzy Sets Syst.*, vol. 461, Jun. 2023, Art. no. 108476.
[32] Z. Yuan, H. Chen, P. Xie, P. Zhang, J. Liu, and T. Li, "Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions," *Appl. Soft Comput.*, vol. 107, Aug. 2021, Art. no. 107353.
[33] D.-D. Zou, Y.-L. Xu, L.-Q. Li, and W.-Z. Wu, "A novel granular variable precision fuzzy rough set model and its application in fuzzy decision system," *Soft Comput.*, vol. 27, no. 13, pp. 8897–8918, Jul. 2023.
[34] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2986–2999, Jul. 2018.
[35] M. Yuan, J. Xu, T. Li, and Y. Sun, "Feature selection based on self-information and entropy measures for incomplete neighborhood decision systems," *Complex Intell. Syst.*, vol. 9, no. 2, pp. 1773–1790, Apr. 2023.
[36] X. Wang, W. Zhang, D. Liu, H. Yu, X. Yang, and X. Yang, "Pseudolabel decision-theoretic rough set," *Math. Problems Eng.*, vol. 2019, Mar. 2019, Art. no. 6810796.
[37] L. Zou, S. Ren, Y. Sun, and X. Yang, "Attribute reduction algorithm of neighborhood rough set based on supervised granulation and its application," *Soft Comput.*, vol. 27, no. 3, pp. 1565–1582, Feb. 2023.
[38] K. Liu, T. Li, X. Yang, H. Ju, X. Yang, and D. Liu, "Feature selection in threes: Neighborhood relevancy, redundancy, and granularity interactivity," *Appl. Soft Comput.*, vol. 146, Oct. 2023, Art. no. 110679.

SHENGLI GAO is currently with the Faculty of Intelligent Engineering Technology, Jiangsu Vocational College of Finance and Economics. His research interests include marching learning, granular computing, and feature selection.