An Emerging Fuzzy Feature Selection Method Using Composite Entropy-Based Uncertainty Measure and Data Distribution

Weihua Xu, Kehua Yuan, Wentao Li, and Weiping Ding

***Abstract*—Feature selection based on neighborhood rough set is a noteworthy step in dealing with numerical data. Information entropy, proven in many theoretical analysis and practical applications, is a compelling feature evaluation method for uncertainty information measures. Nonetheless, information entropy replaces probability with uncertainty measure to evaluate the average amount of information and ignores the decision distribution of data, especially in describing the uncertainty in imbalanced data. This paper discusses an emerging method for the feature selection in fuzzy data with imbalanced data by presenting a local composite entropy based on a neighborhood rough set. Based on the neighborhood rough set model, we discuss a similar relation to describe the relationship between different objects in unbalanced fuzzy data. In this process, to fully consider the distribution characteristics of unbalanced data, we construct a local composite entropy for handling the fuzzy decision systems with uncertainty and decision distribution, which is proven to be monotonic. Moreover, to improve the selection efficiency, a local heuristic forward greedy selection algorithm based on the local composite measure is designed to select the optimal feature subset. Finally, experimental results on twelve public datasets demonstrate that our method has better classification performance than some state-of-the-art feature selection methods in fuzzy data.**

***Index Terms*—Composite information entropy, feature selection, fuzzy decision dataset, local neighborhood rough set.**

I. Introduction

RECENTLY, feature selection, as a required step of data preprocessing, has been widely applied to intelligent computing, data mining, and machine learning [1], [2], [4], [13], [18]. The problems of high-dimensional computation, low classification accuracy and over-fitting could be well solved by removing the irrelevant and redundant features.

Manuscript received 30 November 2021; revised 29 January 2022 and 15 March 2022; accepted 9 April 2022. Date of publication 27 May 2022; date of current version 30 January 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 61976245, 61772002, and 61976120, in part by the Fundamental Research Funds for the Central Universities under Grant SWU119063, in part by the Natural Science Key Foundation of Jiangsu Education Department under Grant 21KJA510004, and in part by the Scientific and Technological Project of Construction of Double City Economic Circle in Chengdu-Chongqing Area under Grant KJCX2020009. *(Corresponding author: Weihua Xu.)*

Weihua Xu, Kehua Yuan, and Wentao Li are with the College of Artificial Intelligence, Southwest University, Chongqing 400715, China (e-mail: chxuwh@gmail.com; yuan945527632@163.com; drliwentao@gmail.com).

Weiping Ding is with the School of Information Science and Technology, Nantong University, Nantong 226019, China (e-mail: dwp9988@163.com).

Digital Object Identifier 10.1109/TETCI.2022.3171784

As a powerful tool to deal with imprecision and uncertainty, rough set theory has attracted extensive attention in feature selection [4], [9], [15], [21]. However, the classical rough set model could only handle classified data, and other types of data need to be converted into discrete data for further processing, which may cause the loss of useful information [24], [26], [31]. Therefore, the rough set model is further extended to deal with various types of data, including set-valued data [14], interval valued data [8], intuitionistic data [7], fuzzy data [5], [12], [15], [22]. Considering the tolerance of neighborhood rough set to data difference, we try to adopt feature selection method based on neighborhood rough set model in fuzzy data. There are many researches on feature selection of neighborhood rough set. Yang *et al.* proposed a novel neighborhood rough set model based on distance metric learning and designed feature selection algorithm according to distance's properties [28]. By fuzzy neighborhood rough sets, Zhang *et al.* proposed a heuristic feature selection algorithm based on fuzzy-neighborhood relative decision entropy [32]. Yang *et al.* studied the dynamic fuzzy neighborhood rough set approach for interval-valued information system with fuzzy decision [29]. Barman *et al.* proposed a novel technique to detect a suitable threshold of neighborhood rough set for hyperspectral band selection [2]. Neighborhood rough set theory is an effective mathematical method for feature selection by depicting uncertainty. It is also noted that the calculation of neighborhood class is a repetitive and redundant process, especially in the traversal selection of features. Therefore, how to improve the efficiency of selection is a problem worth studying.

In order to further improve compute efficiency, Qian *et al.* proposed local neighborhood rough set model and designed the corresponding attribute reduction in [15]. This novel idea only needs to consider the objects in target concept, which provides a convenient way to search the required information directly and reduces the filtering time. It is more effective to search knowledge from local viewpoint than global viewpoint. Because of this remarkable feature of local rough set, scholars focus on several generalized local rough set models for different information systems [3], [5], [11], [26], [34].

It is well known that the applications of uncertainty measures have been reported more and more frequently in feature selection. Wang *et al.* introduced distance measures into fuzzy rough sets and designed attribute significance measure in decision table [23]. Peng *et al.* constructed uncertainty measure based on fuzzy symmetry relations and applied it to feature selection [13]. Wang *et al.* focused on constructing monotonic uncertainty measures in probabilistic rough set model for attribute reduction [22]. The above researches all focus on the uncertainty from algebraic viewpoint and the uncertainty measures could only describe the whole approximate ability of features included in subset. Information entropy can reasonably quantify the statistical characteristics of information, so it and some of its deformations have been widely used in feature selection. Zhang *et al.* investigated the incremental feature selection using a fuzzy-rough-set-based information entropy with incoming instances [33]. Aremu *et al.* proposed a correlation and relative entropy feature engineering framework specific to asset data [1]. Huang *et al.* presented an incremental feature selection method based on the matrix representation of the conditional entropy [6]. It should be pointed out that the importance of features based on information view only explains the impact of uncertainty classification on features. It will be an emerging topic to combine uncertainty measurement with information entropy for feature selection to improve the quality of uncertainty measurement in neighborhood decision system. Sun *et al.* proposed fuzzy neighborhood information entropy based on uncertainty measures to select feature subset in fuzzy neighborhood multigranulation rough sets [18]. Song *et al.* introduced uncertainty measures by using new defined divergence-based cross entropy measure in decision making [17]. Sun *et al.* proposed a novel NMRS-based attribute reduction method using Lebesgue and entropy measures in incomplete neighborhood decision systems [21]. However, the above entropy measures only evaluate the importance of features from the ability of conditional features to deal with uncertainty while ignoring the characteristics of data distribution, which is not applicable to dealing with unbalanced distributed data.

Inspired by the above ideas, we try to improve information entropy to enhance further its ability to describe uncertainty in feature selection. Compared with some existing methods, the designed measure can comprehensively evaluate the features and improve object classification performance. Experimental results also verify the effectiveness of this method. The main contributions of this paper are as follows:

- We improved the ability of information entropy to describe uncertainty and designed a composite information entropy (CIE) to depict features in a fuzzy decision dataset. Compared with some existing methods, this measure could select excellent features that have better object classification performance.
- Unlike other information entropy measures, the designed local composite information entropy focuses on the uncertainty measure and decision distribution, improving its ability to describe uncertainty and better evaluate selection features. Meanwhile, a forward heuristic algorithm can be designed to select essential features based on the monotonicity of the designed measure. The heuristic algorithm settles the NP problem of finding attribute reduction subset very well, significantly reducing the time complexity of selecting the optimal feature subset.
- Moreover, local thought is also considered in the feature selection. The local neighborhood rough set avoids the unnecessary calculation out of target concept, further improving selection efficiency. Compared with the global rough set, the feature selection algorithm based on the local rough set has better classification accuracy and error.

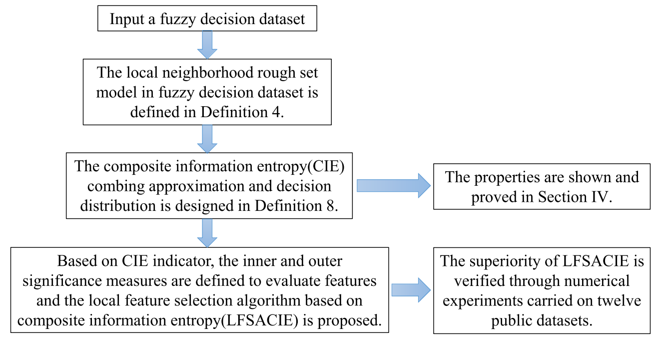


Fig. 1. Block diagram of steps of the proposed approach.

This paper is organized as follows. Some related works about feature selection are introduced in Section II. In Section III, the preliminaries of local neighborhood rough set is reviewed. The composite information entropy and its properties are shown in Section IV. Section V introduces the local heuristic property selection algorithm and analysis its corresponding time complexity. In Section VI, we verify the effectiveness and robustness of designed algorithms in twelve public datasets from the classify accuracy induced by reduction. Finally, the conclusions and the further works are described in Section VII. Moreover, the block diagram of steps of the proposed approach is shown in the Fig. 1.

II. Related Works

With the development of information science, the data scale becomes bigger and the redundant information increases correspondingly. The feature selection as a reduction method has been widely used in data mining. As we all know, the feature evaluation index, selection strategy and stop criteria are the three key elements in feature selection algorithm. These three elements are directly related to the result of feature selection. At present, the researches try to improve the classification performance through designing different select measure according to the characterizes of data.

As an important tool to deal with uncertain information, rough set theory does not need any prior knowledge in conceptual approximation, so it is further applied to feature selection algorithm. For dealing with real data, Qian *et al.* designed the significant measure based on lower approximation to measure the importance of attributes and adopted a forward strategy to select excellent features until the positive region is the same as original ones [26]. Zhou *et al.* applied the rough set theory to online stream feature selection and designed an online early terminated streaming feature selection algorithm. In their proposed algorithm, the dependency degree function is regarded as the early terminated mapping function, and the feature still can be selected unless the function gap between next subset and the current subset is less than a certain value β [35]. Wang *et al.* defined information function based on neighborhood dependency function and designed heuristic feature selection algorithm [27]. As for fuzzy decision data, sheeja *et al.* given the new definition of fuzzy rough set based on diverse measure and designed feature selection algorithm on the corresponding approximations [20]. In 2016, Wang *et al.* defined the dependency between fuzzy decision and conditional attributes and employed the dependency to evaluate the significance of candidate feature, using which a greedy feature subset selection algorithm is designed [4]. Thereupon, they proposed a feature selection algorithm based on fuzzy approximation in parameterized fuzzy neighborhood [25].

To improve the select efficiency, the various strategy is designed. Chen *et al.* utilized the parallel of neighborhood distance matrix to design parallel algorithm for computing reduction in DNRS, the corresponding reduced subset behaves well in classification accuracy and consume time [4]. For dynamic data, Sang *et al.* proposed an incremental feature selection method [16]. Jiang *et al.* defined the intra-class radius and inter-class radius to distinguish samples and also proposed an accelerator attribute reduction algorithm based on supervised neighborhood rough set. Their experimental results show the accelerator reduction algorithm is effective and efficient for feature selection [9].

Feature selection aims to select a minimal feature subset with as much important information as possible. For fuzzy data, we mainly research a feature selection approach to improve the object classification performance. Compared with existing researches, the designed algorithm could evaluate features based on uncertainty measure and data distribution from information entropy viewpoint. All the superiority of proposed algorithm is verified through numerical experiments.

III. Preliminaries

In this section, some basic concepts are reviewed, including fuzzy decision dataset, local neighborhood rough set and some measures of rough set model.

A. Fuzzy Decision Dataset

Definition 1 [26]: Let $FDD = (U, N)$ be a fuzzy decision dataset, where $U = \{x_1, x_2, \ldots, x_n\}$ denotes the set of finite objects, $N = B \cup D$ denotes the union of conditional attribute set $B = \{b_1, b_2, \ldots, b_m\}$, decision attribute set $D = \{d\}$, and $B \cap D = \emptyset$. For $\forall b \in N$ and $\forall x \in U$, $b(x)$ represents the value of object x under the attribute b. Also, if $b(x) \in [0, 1]$ for all conditional attributes, the I is called fuzzy decision dataset. A fuzzy decision dataset is shown in Table I.

TABLE I
A Fuzzy Decision Dataset

U	b_1	b_2	b_3	b_4	d
x_1	0.42	0.74	0.64	0.76	1
x_2	0.39	0.72	0.39	0.75	1
x_3	0.35	0.84	0.34	0.69	1
x_4	0.37	0.79	0.35	0.7	1
x_5	0.54	0.62	0.28	0.77	1
x_6	0.24	0.41	0.31	0.81	1
x_7	0.32	0.42	0.30	0.77	2
x_8	0.15	0.78	0.39	0.35	2
x_9	0.36	0.85	0.45	0.41	2
x_{10}	0.58	0.64	0.43	0.59	2

B. Local Neighborhood Rough Set in Fuzzy Decision Dataset

The classical rough set model can only deal with categorical data accurately due to the strict equivalence relation, which lacks inclusiveness in numerical data. For this limitation, we use the Euclidean distance Δ to describe the relationship between any two objects in universe under neighborhood relation R_A defined on $A \subseteq B$ in a fuzzy decision dataset. Given δ, two samples can be considered to be the same when the distance between them is less than or equal to δ. $\delta_{R_A}(x)$ denotes the neighborhood classes [26] of x on R_A, which is defined as

$$[x]^{\delta}_{R_A} = \{x_s | \Delta(x, x_s) \leq \delta, x_s \in U\}. \quad (1)$$

Based on the neighborhood classes, we will introduce the global and local neighborhood rough set model to approximate target concept by two definite sets, lower approximation and upper approximation.

Definition 2 [26]: Let $FDD = (U, B \cup D)$ be a fuzzy decision dataset, where $U/D = \{D_1, D_2, \ldots, D_s\}$. Given neighbor radius δ, the neighborhood relation and corresponding neighborhood class of object $x \in U$ defined on $A \subseteq B$ can be expresses as R^{δ}_A and $[x]^{\delta}_{R_A}$. Then, for any $D_i \in U/D$, the global lower and upper approximations of D_i under R_A are respectively defined as

$$\begin{aligned} \underline{R^{\delta}_{G,A}}(D_i) &= \left\{x | [x]^{\delta}_{R_A} \subseteq D_i, x \in U\right\} \\ &= \left\{x | [x]^{\delta}_{R_A} \subseteq D_i, x \in D_i\right\}, \\ \overline{R^{\delta}_{G,A}}(D_i) &= \left\{x | [x]^{\delta}_{R_A} \cap D_i \neq \emptyset, x \in U\right\} \\ &= \cup \left\{[x]^{\delta}_{R_A} | [x]^{\delta}_{R_A} \cap D_i \neq \emptyset, x \in D_i\right\}. \end{aligned} \quad (2)$$

The pair $< \underline{R^{\delta}_{G,A}}(D_i), \overline{R^{\delta}_{G,A}}(D_i) >$ is called global neighborhood rough set.

According to Definition 2, We need to compare the relationship between target concept and the object class in U. It is a time-consuming project for obtaining the target approximations by traversing all objects, especially in big scale data. Actually, the target objects what we looking for can be obtained through the target concept. The local rough set model proposed by Qian *et al.* has settled this problem well.

Definition 3 (Continue to Definition III-B): For any $D_i \in U/D$, the local lower and upper approximations of D_i under R_A are respectively defined as

$$\begin{aligned} \underline{R^{\delta}_{A}}(D_i) &= \left\{x | [x]^{\delta}_{R_A} \subseteq D_i, x \in D_i\right\}, \\ \overline{R^{\delta}_{A}}(D_i) &= \left\{x | [x]^{\delta}_{R_A} \cap D_i \neq \emptyset, x \in D_i\right\}. \end{aligned} \quad (3)$$

The pair $< \underline{R_A^\delta}(D_i), \overline{R_A^\delta}(D_i) >$ is called local neighborhood rough set. The local rough set only needs to compare the relationship between target concept and object classed in D_i while the global rough set needs to obtain the all information from the universe U according to definition III-B, thus it could reduce the time complexity in concept approximation.

Definition 4 (Continue to Definition III-B): The local lower approximation $\underline{R_A^\delta}(D)$ and upper approximation $\overline{R_A^\delta}(D)$ in fuzzy decision dataset are defined by

$$\underline{R_A^\delta}(D) = \left\{\underline{R_A^\delta}(D_1), \underline{R_A^\delta}(D_2), \ldots, \underline{R_A^\delta}(D_s)\right\},$$
$$\overline{R_A^\delta}(D) = \left\{\overline{R_A^\delta}(D_1), \overline{R_A^\delta}(D_2), \ldots, \overline{R_A^\delta}(D_s)\right\}. \quad (4)$$

The local positive, negative and boundary regions are $POS_{R_A^\delta}(D) = \bigcup_{i=1}^s \underline{R_A^\delta}(D_i)$, $NEG_{R_A^\delta}(D) = \bigcup_{i=1}^s (U - \overline{R_A^\delta}(D_i)$ and $BND_{R_A^\delta}(D) = \bigcup_{i=1}^s (\overline{R_A^\delta}(D_i) - \underline{R_A^\delta}(D_i))$.

C. Accuracy Measures of Rough Set Model

In neighborhood rough set theory, the accuracy of model depends on the size of positive, negative and boundary regions [26]. The smaller the positive region and the larger the negative region are, the weaker its accuracy is. Moreover, the smaller the boundary region of rough model is, the stronger ability to deal with uncertainty is. The following measures are usually used to describe the approximate ability of rough set.

Definition 5 [30]: Given a fuzzy decision dataset $FDD = (U, N)$, the neighborhood radius is δ. $\forall D_i \in U/D$, the accuracy of approximations

$\gamma_{R_A^\delta}(D_i)$ in neighborhood rough set is defined as

$$\gamma_{R_A^\delta}(D_i) = \frac{|\underline{R_A^\delta}(D_i)|}{|D_i|}. \quad (5)$$

In order to further consider the possibility of the upper approximation, the accuracy of approximations $\alpha(R_A^\delta)(D_i)$ is also defined as

$$\alpha_{R_A^\delta}(D_i) = \frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|}. \quad (6)$$

The accuracy of approximation $\alpha_{R_A^\delta}(D_i)$ could comprehensive describe the accuracy of rough set model from certain and possible aspects. The above two measures depict neighborhood rough set based on approximations, they all have a positive relationship with the approximate ability, that is, the larger the value, the more accurate the model.

The uncertainty could be described by the amount of information, and the greater the amount of information, the less the uncertainty. In 1948, Shannon put forward the concept of Shannon entropy to describe the amount of information, which is defined as follows.

Definition 6 [19]: For random variance Y, where its value range is $\{Y_1, Y_2, \ldots, Y_s\}$, $p(Y_i)$ is the corresponding probability of Y_i, then the shannon information entropy $H(p)$ is

$$H(p) = -\sum_{Y_i \in Y} p(Y_i) log(p(Y_i)). \quad (7)$$

D. Motivation

In data analysis, removing irrelevant and redundant features is a necessary step for further processing. Recently, rough set model as an important tool to deal with uncertainty knowledge has been widely used to feature selection. However, most of the existing studies focus on uncertainty measures to evaluate features, but ignore the distribution of data. Therefore, we try to design an indicator combining uncertainty and data distribution to evaluate feature from the perspective of information entropy. Compared with other existing indicators, the indicator proposed in this paper can describe characteristics more comprehensively, as shown in the following Example.

Example 1: A fuzzy decision dataset is shown in Table I, where $U = \{x_1, x_2, \ldots, x_{10}\}$ and B contains four conditional attributes $\{b_1, b_2, b_3, b_4\}$ and d is decision attribute.

We divide the conditional features into two granules, that is, $B_1 = \{b_1, b_2\}$ and $B_2 = \{b_3, b_4\}$. For obtaining the lower and upper approximations of two decision classes under B_1, we first calculate the distance between any two objects under B_1, which is shown in distance matrix D_1.

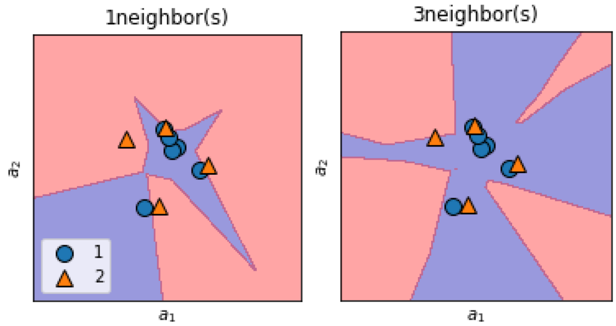
$$M_1 = \begin{bmatrix} 0 & 0.04 & 0.12 & 0.07 & 0.17 & 0.38 & 0.34 & 0.27 & 0.13 & 0.19 \\ 0.04 & 0 & 0.13 & 0.07 & 0.18 & 0.34 & 0.31 & 0.25 & 0.13 & 0.21 \\ 0.12 & 0.13 & 0 & 0.05 & 0.29 & 0.44 & 0.42 & 0.21 & 0.01 & 0.30 \\ 0.07 & 0.07 & 0.05 & 0 & 0.24 & 0.40 & 0.37 & 0.22 & 0.06 & 0.26 \\ 0.17 & 0.18 & 0.29 & 0.24 & 0 & 0.37 & 0.30 & 0.42 & 0.29 & 0.04 \\ 0.38 & 0.34 & 0.44 & 0.40 & 0.37 & 0 & 0.08 & 0.38 & 0.46 & 0.41 \\ 0.34 & 0.31 & 0.42 & 0.37 & 0.30 & 0.08 & 0 & 0.40 & 0.43 & 0.34 \\ 0.27 & 0.25 & 0.21 & 0.22 & 0.42 & 0.38 & 0.40 & 0 & 0.22 & 0.45 \\ 0.13 & 0.13 & 0.01 & 0.06 & 0.29 & 0.46 & 0.43 & 0.22 & 0 & 0.30 \\ 0.19 & 0.21 & 0.30 & 0.26 & 0.04 & 0.41 & 0.34 & 0.45 & 0.30 & 0 \end{bmatrix}.$$

Let $\delta = 0.16$, the neighborhood on relation $R_{\{B_1\}}$ are
$[x_1]_{R_{B_1}}^\delta = [x_2]_{R_{B_1}}^\delta = [x_3]_{R_{B_1}}^\delta = [x_4]_{R_{B_1}}^\delta = [x_9]_{R_{B_1}}^\delta = \{x_1, x_2, x_3, x_4, x_9\}$,
$[x_5]_{R_{B_1}}^\delta = [x_{10}]_{R_{B_1}}^\delta = \{x_5, x_{10}\}$,
$[x_6]_{R_{B_1}}^\delta = [x_7]_{R_{B_1}}^\delta = \{x_6, x_7\}$,
$[x_8]_{R_{B_1}}^\delta = \{x_8\}$.

Then, we obtain the lower and upper approximations of $D_i (i = 1, 2)$ according to Definition III-B as follows.
$\underline{R_{\{B_1\}}^\delta}(D_1) = \emptyset$,
$\underline{R_{\{B_1\}}^\delta}(D_2) = \{x_8\}$;
$\overline{R_{\{B_1\}}^\delta}(D_1) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$,
$\overline{R_{\{B_1\}}^\delta}(D_2) = \{x_7, x_8, x_9, x_{10}\}$.

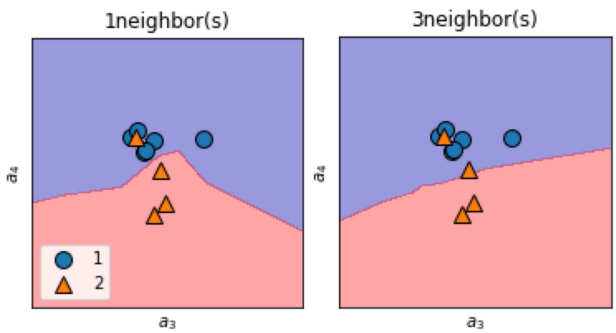
Then, the $\alpha_{R_{B_1}^\delta}(D_i)$ are obtained
$\alpha_{R_{B_1}^\delta}(D_1) = \frac{1}{10}$, $\alpha_{R_{B_1}^\delta}(D_2) = \frac{1}{10}$.

It is noted that the $\alpha_{R_A^\delta}(D)$ is set to $\frac{1}{n}$ for convenience when the lower approximation is $\emptyset$. According to the lower approximation, there is no object that can be determined to belong to D_1, and there is only one object that is determined



(a) (b)

Fig. 2. The classification results of KNN on B_1.



(a) (b)

Fig. 3. The classification results of KNN on B_2.

Algorithm 1: Local Feature Selection Algorithm Based on Composite Information Entropy(LFSACIE).

Input :
1) A fuzzy decision dataset $FDD = (U, B \cup D)$, where $U/D = \{D_1, D_2, \ldots, D_s\}$;
2) The radius of neighborhood δ.

Output : Feature subset A.

begin
- 1:set $A \leftarrow \emptyset$, $CIE(\emptyset, D) \leftarrow 0$.
- 2:compute $CIE(B, D)$ of feature set B;
- 3:compute $CIE(B - \{b\}, D)$ for $b \in B$;
- 4:select feature $b_o = \underset{b \in B}{argmax}\, IM(b, B, D)$;
- 5:$A \leftarrow b_o$;
- 6:compute $CIE(\{b\}, D), start = 1$;
- 7:**while** *start=1* **do**
  - 8:**for** *each* $b \in B - A$ **do**
    - **for** $i = 1 : s$ **do**
      - compute $\underline{R^{\delta}_{A\cup\{b\}}}(D_i), \overline{R^{\delta}_{A\cup\{b\}}}(D_i)$;
    - **end**
    - compute $CIE(A \cup \{b\}, D), SM(b, A, D)$;
  - **end**
  - 9:select $b_k = \underset{b \in B-A}{argmax}\, SM(b, A, D)$;
  - 10:**if** $SM(b_k, A, D) \geq \beta$ **then**
    - $A = A \cup \{b_k\}$,
    - $start = 1$;
  - **else**
    - $start = 0$;
  - **end**
- **end**
- 11:**return** : A.

end

to belong to D_2. Similarly, we could obtain the lower and upper approximations of two decision classes under B_2 are

$\underline{R^{\delta}_{\{B_2\}}}(D_1) = \{x_1\}$,

$\underline{R^{\delta}_{\{B_2\}}}(D_2) = \{x_8, x_9\}$;

$\overline{R^{\delta}_{\{B_2\}}}(D_1) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$,

$\overline{R^{\delta}_{\{B_2\}}}(D_2) = \{x_7, x_8, x_9, x_{10}\}$.

Then, the $\alpha_{R^{\delta}_{B_2}}(D_i)$ are

$\alpha_{R^{\delta}_{B_2}}(D_1) = \frac{1}{10}$, $\alpha_{R^{\delta}_{B_2}}(D_2) = \frac{1}{5}$.

The information entropy index combing with lower and upper approximation reflects the uncertain information on different classes, and it has a reverse relation with certainty. The values of information entropy $I_A(D) = \sum_{D_i \in U/D} \alpha_{R^{\delta}_A}(D_i) log \alpha_{R^{\delta}_A}(D_i)$ under B_1 and B_2 are 0.58 and 0.65, respectively. Thus, the feature subset B_1 is considered to be better than B_2. The classification results of 1NN and 3NN on B_1 and B_2 are shown in Fig. 2 and Fig. 3.

The subgraph (a) and (b) in Fig. 2 characterise the predict classification results on B_1 when the neighbor parameter is 1 and 3, respectively. From these two pictures, we could find that the decision boundary is not clear, and none of the cases in the second category are classified correctly. However, we could find that the decision boundary of K-Nearest Neighbors(KNN) classifier on B_2 is clear, and there is only an object that is clearly misclassified. From the classification comparison between B_1 and B_2, the feature subset B_2 stands out more than B_1.

According to above analysis, we know that the feature selection based on the indicator combing approximation space and decision distribution could achieve the better classification. In this paper, we further design feature selection algorithm combing with data class distribution. The innovations of this paper are as follows: (1) Introduce composite information entropy to depict the importance of feature in neighborhood rough set. (2) Compared with global rough set, the local rough set only needs to obtain the object information in target decision, which avoids some unnecessary computation. In order to improve reduction efficiency, we select feature from the local viewpoint inspired by this idea. (3) A greedy heuristic forward algorithm based on composite information measure is designed for feature selection.

IV. Composite Information Entropy Measure Based on Local Neighborhood Rough Set in Fuzzy Decision Dataset

In a decision dataset, the rough set model is employed for depicting uncertainty caused by the difference of lower approximation and upper approximation, and this uncertainty measure only describes the feature contained in a subset of features. Meanwhile, there are many class imbalanced data, and the distribution of these data will affect the evaluation of features. Therefore, we propose an emerging composite information entropy to measure rough set model combing with approximations and all kinds of distribution inspired by Shannon entropy.

Definition 7: Given $FDD = (U, B \cup D)$. Let neighborhood radius δ, the $\underline{R_A^\delta}(D)$ and $\overline{R_A^\delta}(D)$ denote local lower and upper approximation on neighborhood relation R_A. Then the composite information entropy $CIE(A, D)$ is defined as

$$CIE(A, D) = -\sum_{i=1}^{s} \frac{|D_i|}{|U|} log \frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|}. \quad (8)$$

The composite information entropy measure depicts the uncertainty more detail due to the combination of $\frac{|D_i|}{|U|}(i = 1, 2, \ldots, s)$ and approximations. It is worth noting that the local upper approximation of the decision target is itself according to the definition of approximations, thus $\frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|}$ is reduced to $\frac{|\underline{R_A^\delta}(D_i)|}{|D_i|}$. This measure satisfies the following properties.

Proposition 1: For $\forall A \subseteq B, CIE(A, D) \geq 0$.

Proof: According to the definition of rough set, we know $\frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|} \in [0, 1]$ for all target decisions, $log \frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|} \leq 0$ for $i = 1, 2, \ldots, s$, then $CIE(A, D) \geq 0$.

Proposition 2: For $\forall A \subseteq B$,

1) If there exists D_i satisfying $\frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|} = 0 (i = 1, 2, \ldots, s)$, then $CIE(A, D) \to \infty$;

2) If $\frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|} = 1$ for all $D_i (i = 1, 2, \ldots, s)$, then $CIE(A, D) = 0$.

Proof: The above two properties are easily obtained according to the definitions of rough set and $CIE(A, D)$.

Proposition 3: Let $A \subseteq A'$, then $CIE(A', D) \leq CIE(A, D)$.

Proof: Given a certain δ, because $A \subseteq A'$, $R_{A'} \subseteq R_A$, then $\underline{R_A^\delta}(D_i) \subseteq \underline{R_{A'}^\delta}(D_i)$ and $\overline{R_{A'}^\delta}(D_i) \subseteq \overline{R_A^\delta}(D_i)$, further we have $\frac{|\underline{R_{A'}^\delta}(D_i)|}{|\overline{R_{A'}^\delta}(D_i)|} \geq \frac{|\underline{R_A^\delta}(D_i)|}{|\overline{R_A^\delta}(D_i)|}$ for $i = 1, 2, \ldots, s$. It is easily to be obtained that $CIE(A', D) \leq CIE(A, D)$ according to the properties of log function.

There is a reverse relationship between $CIE(A, D)$ and neighborhood relation R, that is, with the finer the relationship, the higher the approximation ability of the rough set model is, and the smaller the value of $CIE(A, D)$ is. When the $\underline{R_{A'}^\delta}(D_i)| = |\overline{R_{A'}^\delta}(D_i)|$ for all target decisions, the $CIE(A, D)$ reaches the minimum 0.

Example 2: Continue to Example 2.1, we further compute the composite information entropy $CIE(\{B_i\}, D_j)(i, j = 1, 2)$ to measure the importance of features combined with approximations and decision distribution.

The decision distribution is

$P(D_1) = \frac{3}{5}$, $P(D_2) = \frac{2}{5}$.

Thus, the value of composite information entropy $CIE(\{B_i\}, D)(i = 1, 2)$ are

$CIE(\{B_1\}, D) = 1.94$, $CIE(\{B_2\}, D) = 1.35$.

The smaller the value of it, the less uncertainty information in the neighborhood, then the attribute B_2 will be selected according to CIE. The composite information entropy has more advantages in the process of feature selection.

Definition 8: Given $FDD = (U, B \cup D)$ and $A \subseteq B$, then A is a feature select reduction of B iff:

1) $CIE(A, D) = CIE(B, D)$;

2) $CIE(A', D) \nleq CIE(B, D)$ for any $A' \subset A$.

V. Feature Selection Algorithm Based on Composite Information Entropy Measure

The composite information entropy, combing with approximations of target concept and distribution of decision class, could comprehensively reflect the approximation ability of rough set model defined on a certain feature subset.

1) The approximate ability could reflected by the value of $CIE(A, D)$. The higher the value of CIE, then the weaker the ability of A to deal with uncertainty.

2) With the increase of the number of features, the higher the accuracy of rough set model, and the smaller the value of $CIE(A, D)$ is.

Given a feature subset $A \subseteq B$, we could measure the importance of attribute a respect to A through the value of $CIE(A \cup \{a\}, D)$. The smaller the value of $CIE(A \cup \{a\})$ is, the greater the reduced uncertainty is, that is, the more important the attribute is in the process of approximation.

From Definition 9, we know that there may have multiple reduction sets, but one attribute reduction is enough in some cases. As for this problem, we choose a heuristic forward greedy attribute selection algorithm to select features that have the same approximate ability with original data. We will firstly define two significance measures to depict feature as follows.

Definition 9: Given $FDD = (U, B \cup D)$ and neighborhood radius δ. For $\forall b \in B$, the inner significance measure of b with respect to B is defined as

$$IM(b, B, D) = CIE(B - \{b\}, D) - CIE(B, D). \quad (9)$$

The higher the value of $IM(b, B, D)$, the higher the increase of information entropy relative to $CIE(B, D)$, which indicates that feature b is the most important relative to B. Therefore, we first select the feature b satisfying $argmax_{b \in B} IM(b, B, D)$.

Definition 10: Given $FDD = (U, B \cup D)$ and neighborhood radius δ. Let $A \subseteq B$, for $\forall a \in B - A$, the outer significance measure of b with respect to A is defined as

$$SM(b, A, D) = CIE(A, D) - CIE(A \cup \{b\}, D). \quad (10)$$

The higher the value of $SM(b, A, D)$, the more it decreases relative to $CIE(A, D)$, which illustrates the importance of feature b respect to A.

In order to improve the efficiency of selection, a heuristic forward algorithm is adopted to choose the important feature subset in fuzzy decision datasest. We will confirm the first selected feature b_o satisfying $argmax_{b \in B} IM(b, B, D)$, and then select the other excellent ones according to the maximum principle of indicator SM. It is worth noting that the larger SM is, the more important b is relative to A. When the $SM(b, A, D) \to 0$, the b could not be selected. Let $\beta = 0.05$, the process terminates when

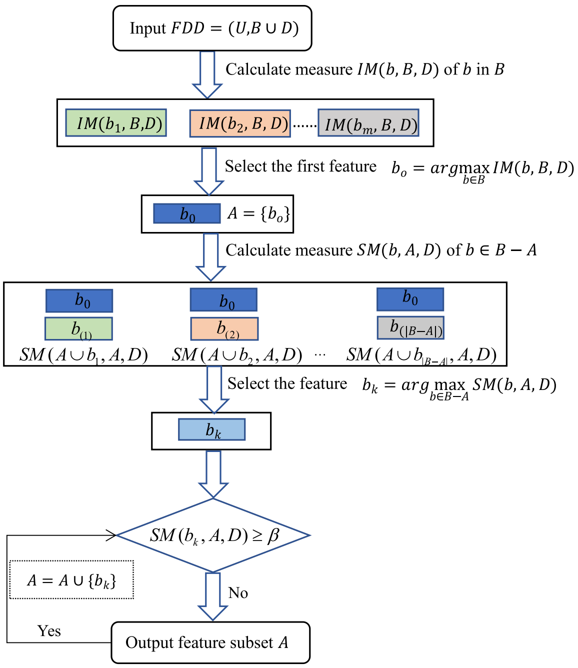


Fig. 4. The process of selecting excellent features based on composite information entropy.

the judgment condition is not satisfied. The more detail selection process and pseudo code is shown in Fig. 4 and **Algorithm 1**, respectively.

On the basis of measures SM and IM, the local feature selection algorithm based on composite information entropy (LFSACIE) is designed and shown in **Algorithm 1**. Similarly, the corresponding feature selection algorithm based on global rough set is called as GFSACIE. In step 2, we first need to compute the local lower and upper approximations according to neighborhood in target decision, whose time complexity is $O(\sum_{j=1}^{s}((|B|+1)|D_j||U|+2|D_j|^2))$, then we obtain the composite information entropy $CIE(B,D)$ whose time complexity is $O(s)$, thus the complexity of step 2 is $O(\sum_{j=1}^{s}((|B|+1)|D_j||U|+2|D_j|^2+1)$ while that of global rough set is $O((|B|+1)|U|^2 + 2\sum_{j=1}^{s}(|D_j||U|+1))$. Similarly, the time complexity of step 3 is $O(|B|(\sum_{j=1}^{s}(|B||D_j||U|+2|D_j|^2+1)))$. If a_o is the i_{th} selected feature, the time complexity of step 8 is $O((|B|-i+1)(\sum_{j=1}^{s}((i+1)|D_j||U|+2|D_j|^2+1)+1))$ due to the computation of CIE and SM. In the process of obtaining optimal feature a_o, we need to select the feature satisfying $argmax_{a_k \in B-A} SM(a_k,A,D)$, whose time complexity is $O(|B|-i+1)$. Also, the complexity is $O(1)$ for step 10. Suppose the reduction set contains l properties finally, the whole time complexity of step 7 is $O(\sum_{i=1}^{l}(|B|-i+1)(\sum_{j=1}^{s}((i+1)|D_j||U|+2|D_j|^2+1))$ $+1+|B|-i)$ while that of global one is $O(\sum_{i=1}^{l}(|B|-i+1)((i+1)|U|^2+\sum_{j=1}^{s}(2|D_j||U|+1)+1+|B|-i))$ because the corresponding rough set model needs to obtain all information granular determined by objects from whole universe. Due to $|D_j||U| \ll |D_j|^2$, the local algorithm could reduce the compute complexity compared with global algorithm, especially in Big Data.

TABLE II
DATASETS DESCRIPTION

Nos.	Datasets	Objects	Attributes	Classes	Type
set1	Hcv	615	13	4	Real
set2	Audit Data	772	18	2	Real
set3	Breast Cancer Coimbra	116	10	2	Real
set4	Chemical Composionof Ceramic	88	18	2	Real
set5	Climate Model Simulation Crashes	504	18	2	Real
set6	Hill	606	101	2	Real
set7	Ionosphere	351	34	2	Real
set8	Speaker Accent Recognition DataSet	329	13	6	Real
set9	Wdbc	569	31	2	Real
set10	Parkinsons	195	23	2	Real
set11	Wine	178	14	3	Real
set12	Appendicitis	106	8	2	Real

TABLE III
THE NUMBER OF FEATURES SELECTED BY LFSACIE UNDER DIFFERENT RADIUS ON TWELVE DATASETS

Datasets	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
set1	3	3	3	4	4	5	5	5	5	6
set2	2	2	2	2	2	2	5	6	6	6
set3	2	3	3	3	4	4	4	4	4	5
set4	2	2	2	2	2	2	2	2	2	2
set5	2	3	3	3	3	3	3	4	4	4
set6	6	6	8	9	10	13	11	10	10	10
set7	4	4	4	5	5	5	5	5	5	6
set8	3	3	4	4	5	5	6	6	6	6
set9	2	3	3	3	4	4	4	4	5	5
set10	2	2	3	3	3	4	4	4	4	4
set11	2	2	3	3	3	3	3	4	4	4
set12	2	2	2	3	3	3	4	4	4	4

VI. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we design a series of numerical experiments to verify the effectiveness of the proposed algorithm (LFSACIE), which is mainly reflected in the following three aspects: (1) the number of selected attributes; (2) the superiority of local feature selection algorithm compared with global algorithm and (3) the accuracy performance of selection algorithms under different classifiers.

A. Experimental Design

Three attribute reduction methods based on rough set theory are selected to compare with the LFSACIE, and the more details about them are as follows.

1) A feature selection algorithm based on fuzzy rough set(FSAFRS) [20]: Sheeja *et al.* introduced a new fuzzy rough sets based on the divergence measure of fuzzy sets, and then designed a feature selection algorithm using fuzzy positive region.

TABLE IV
THE CLASSIFICATION PERFORMANCE OF FEATURES SELECTED BY LFSACIE UNDER DIFFERENT RADIUS ON TWELVE DATASETS

Datasets	Classifiers	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
set1	DT	75.44(4.52)	74.49(5.17)	73.33(7.48)	**88.13(3.79)**	75.95(5.27)	83.59(4.8)	84.06(3.84)	85.2(3.71)	83.1(6.36)	84.72(4.83)
	KNN	89.92(2.84)	89.75(2.99)	89.92(3.51)	**90.72(3.54)**	90.1(3.58)	89.61(4.21)	89.43(2.18)	89.6(4.19)	89.26(5.17)	89.27(4.23)
	Average	82.68(3.68)	82.12(4.08)	81.62(5.49)	**89.43(3.67)**	83.02(4.43)	86.6(4.51)	86.75(3.01)	87.4(3.95)	86.18(5.77)	87(4.53)
set2	DT	**100(0)**	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)
	KNN	**100(0)**	99.74(0.82)	100(0.82)	99.74(0.82)	99.74(0.82)	99.61(0)	98.44(1.74)	98.32(1.33)	98.45(1.34)	98.71(1.71)
	Average	**100(0)**	99.87(0.41)	100(0.41)	99.87(0.41)	99.87(0.41)	99.81(0)	99.22(0.87)	99.16(0.67)	99.22(0.67)	99.35(0.86)
set3	DT	43.03(8.66)	41.29(11.51)	50.68(11.28)	36.14(14.38)	**56.89(11.48)**	40.53(13.15)	34.7(15.46)	38.94(22.06)	35.91(13.34)	47.05(17.12)
	KNN	68.03(10.46)	56.67(15.41)	66.59(13.22)	64.47(10.96)	**72.5(10.89)**	61.44(18.47)	58.64(15.08)	56.97(13.31)	60.15(11.36)	61.89(14.32)
	Average	55.53(9.56)	48.98(13.46)	58.64(12.25)	50.3(12.67)	**64.7(11.18)**	50.98(15.81)	46.67(15.27)	47.95(17.69)	48.03(12.35)	54.47(15.72)
set4	DT	**100(0)**	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)
	KNN	**100(0)**	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)
	Average	**100(0)**	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)	100(0)
set5	DT	81.48(5.98)	83.7(5.15)	83.89(4.79)	82.04(5.59)	83.33(4.28)	83.89(4.1)	83.33(5.72)	88.33(4.28)	**89.07(4.14)**	84.81(6.04)
	KNN	91.48(1.99)	90.93(3.08)	91.48(3.83)	91.67(2.51)	92.04(4.71)	91.85(3.05)	90.93(4.4)	**95(2.76)**	93.89(2.15)	92.22(2.73)
	Average	86.48(3.99)	87.31(4.12)	87.69(4.31)	86.85(4.05)	87.69(4.49)	87.87(3.57)	87.13(5.06)	**91.67(3.52)**	91.48(3.14)	88.52(4.38)
set6	DT	37.94(8.21)	37.63(4.81)	38.78(7.77)	39.28(7.4)	32.01(4.01)	43.24(5.32)	43.9(7.66)	**44.72(7.94)**	41.58(8.36)	39.3(6.71)
	KNN	50.64(6.62)	52.15(7.25)	51.03(9.21)	51.49(4.77)	46.71(6.47)	52.28(4.91)	51.16(5.74)	**55.28(7.02)**	50.66(5.87)	51.51(6.49)
	Average	44.29(7.42)	44.89(6.03)	44.9(8.49)	45.39(6.08)	39.36(5.24)	47.76(5.12)	47.53(6.7)	**50(7.48)**	46.12(7.11)	45.4(6.6)
set7	DT	79.77(3.69)	71.79(6.27)	73.19(8.06)	76.38(9.32)	80.6(5.89)	80.07(7.52)	75.79(6.34)	76.35(3.34)	78.06(7.76)	**80.9(7.77)**
	KNN	91.17(2.81)	85.17(5.7)	87.76(4.21)	87.46(5.26)	87.44(6.36)	**92.05(5.14)**	89.17(2.65)	86.63(4.59)	89.47(4.44)	90.31(4.71)
	Average	85.47(3.25)	78.48(5.99)	80.48(6.13)	81.92(7.29)	84.02(6.12)	**86.06(6.33)**	82.48(4.49)	81.49(3.96)	83.76(6.1)	85.6(6.24)
set8	DT	**51.06(6.6)**	30.69(6.26)	31.58(8.63)	30.09(9.21)	30.09(8.98)	33.13(11.62)	43.19(10.44)	46.84(14.13)	44.68(4.48)	41.93(5.57)
	KNN	**71.15(5.76)**	50.46(4.85)	54.11(9.48)	58.06(6.78)	57.1(7.39)	63.23(5.74)	70.22(8.4)	69.02(10.61)	68.12(7.76)	70.17(7.52)
	Average	**61.1(6.18)**	40.58(5.55)	42.85(9.05)	44.08(7.99)	43.6(8.19)	48.18(8.68)	56.7(9.42)	57.93(12.37)	56.4(6.12)	56.05(6.54)
set9	DT	86.81(3.15)	87.52(2.8)	**90.68(3.63)**	89.81(4.59)	89.63(3.74)	88.23(5.6)	90.34(5.66)	89.63(3.15)	89.64(3.99)	90.15(3.55)
	KNN	93.31(3.01)	95.43(2.89)	95.08(4.12)	95.25(2.49)	94.55(2.81)	95.26(2.75)	**95.6(2.59)**	95.08(2.97)	94.38(1.61)	94.9(2.1)
	Average	90.06(3.08)	91.48(2.84)	92.88(3.87)	92.53(3.54)	92.09(3.27)	91.75(4.17)	**92.97(4.13)**	92.35(3.06)	92.01(2.8)	92.53(2.82)
set10	DT	74.45(8.42)	71.39(13.74)	76.45(5.81)	78.37(12.04)	75.5(10.05)	**78.5(7.22)**	75.97(10.8)	76.47(13.74)	71.34(6.11)	70.71(10.76)
	KNN	86.29(12.62)	86.68(6.78)	90.89(7.73)	90.21(7.58)	90.26(6.98)	**93.84(6.21)**	91.79(5.58)	90.76(7.56)	91.34(5.85)	90.79(5.27)
	Average	80.37(10.52)	79.04(10.26)	83.67(6.77)	84.29(9.81)	82.88(8.52)	**86.17(6.71)**	83.88(8.19)	83.62(10.65)	81.34(5.98)	80.75(8.02)
set11	DT	73.59(9.45)	**93.27(10.43)**	80.33(5.59)	83.07(10.48)	79.87(10.58)	83.2(6.36)	83.1(6.56)	90.52(9.08)	91.47(6.89)	88.82(10.48)
	KNN	88.76(5.25)	**95.56(7.6)**	90.98(4.76)	89.25(10.79)	89.38(8.47)	88.82(7.41)	89.31(4.97)	95.52(4.39)	94.28(9.15)	93.82(6.65)
	Average	81.18(7.35)	**94.41(9.02)**	85.65(5.18)	86.16(10.63)	84.62(9.53)	86.01(6.88)	86.21(5.77)	93.02(6.73)	92.88(8.02)	91.32(8.56)
set12	DT	73.73(11.4)	71.82(12.51)	74.09(19.43)	75.36(16.04)	72.82(16.16)	67.82(21.27)	**76.27(12.24)**	74.73(12.25)	72.27(16.17)	69(14.24)
	KNN	84.18(11.5)	83.91(11.67)	83.82(11.31)	86.64(8.32)	87.73(9.9)	**88.82(8.53)**	88.82(10.81)	86.91(8.85)	85.73(9.54)	86.91(7.74)
	Average	78.95(11.45)	77.86(12.09)	78.95(15.37)	81(12.18)	80.27(13.03)	78.32(14.9)	**82.55(11.52)**	80.82(10.55)	79(12.86)	77.95(10.99)

2) A novel early terminated online streaming feature selection framework(OSFS-ET) [35]: Zhou *et al.* choose the dependency degree function in rough set theory as the early terminated mapping function to construct online streaming feature selection, where $K = 9$ about neighbors and $\beta = 0.01$ for select criteria.

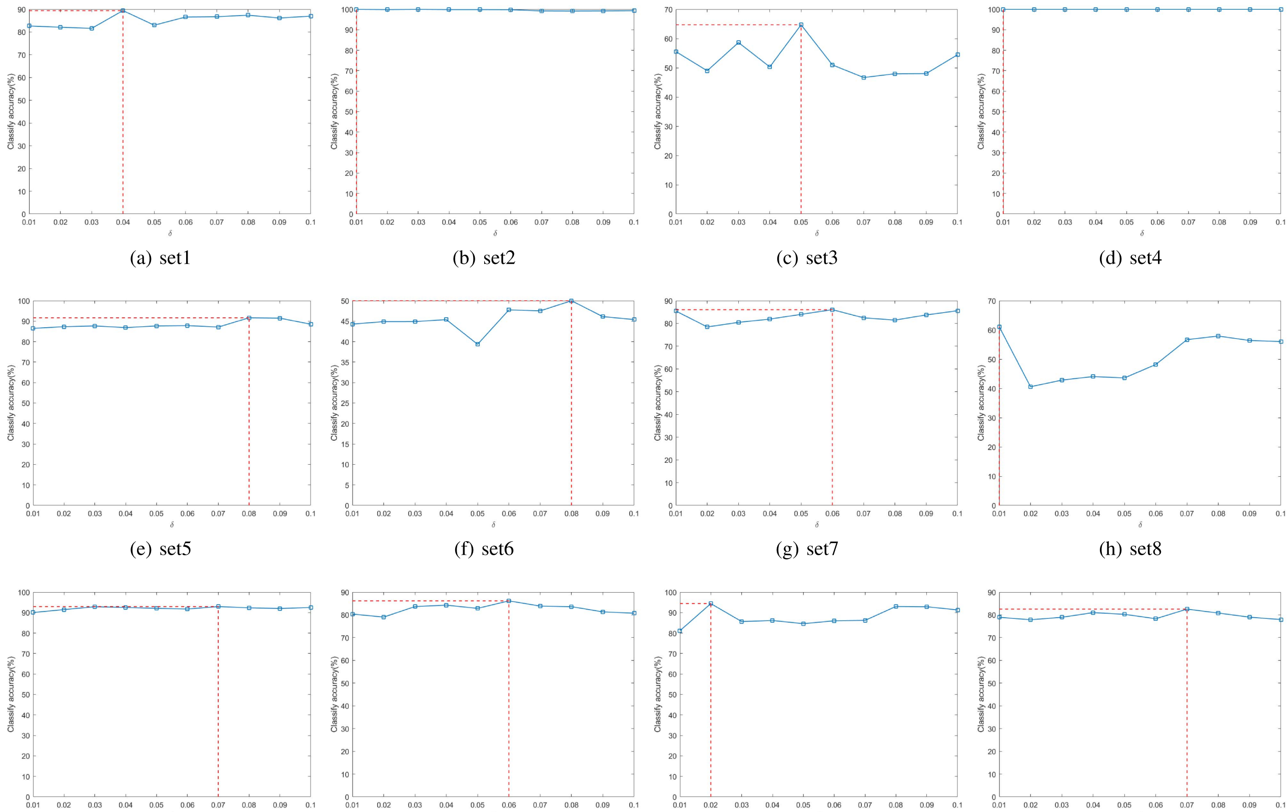
3) A local attribute reduction algorithm proposed by Qian *et al.* (LARD) [26]: the core idea is to keep the certainty of the rough set model no less than the original model. The feature selection measure is the same as LFSALA, and the neighborhood radius $\delta = 0.001$.

4) A local feature selection algorithm based on self-information (LFSASI): Wang *et al.* have explored some measures to describes the importance of attribute and proposed corresponding feature selection mechanism [24]. In the process of attribute selection, $I_B(D)$ shows high distinguishing ability of feature subset B. Thus, they adopted this measure to design heuristic reduction algorithm to compare with LFSACIE under the same conditions, where $\delta = 0.15$, $\beta = 0.01$.

5) Feature subset selection based on fuzzy neighborhood rough set(FSFNRS) [27]: To make the new model tolerate

TABLE V
THE CONSUME TIME OF GFSACIE AND LFSACIE

Time	set1	set2	set3	set4	set5	set6
GFSACIE	11.29	33.17	0.78	2.19	11.00	645.38
LFSACIE	0.93	4.00	0.05	0.64	1.32	99.03
Time	**set7**	**set8**	**set9**	**set10**	**set11**	**set12**
GFSACIE	21.80	6.25	33.81	5.41	2.25	0.34
LFSACIE	3.37	0.25	5.72	0.81	0.19	0.04



(a) set1 (b) set2 (c) set3 (d) set4 (e) set5 (f) set6 (g) set7 (h) set8 (i) set9 (j) set10 (k) set11 (l) set12

Fig. 5. The average classify performance on different feature subset induced by different neighborhood radius.

noises in data, Wang *et al.* built a variable-precision fuzzy neighborhood rough set and designed corresponding feature selection algorithm based on dependency measure constructed on lower approximation, where $\lambda = 0.2, \alpha = 0.9$.

6) A fitting model for feature selection with fuzzy rough sets(FMFSFRS) [25]: Wang *et al.* first constructed a new fuzzy rough set model through introducing a parameter, then defined the significance measure of a candidate attribute based on fuzzy dependency function and designed a greedy forward algorithm for feature selection, where $\lambda = 0.1, \delta = 0.3$.

7) Factor Analysis: As a feature dimension method, it can find hidden representative factors in many variables. Classifying variables of the same nature into one factor can reduce the number of variables and test the hypothesis of the relationship between variables.

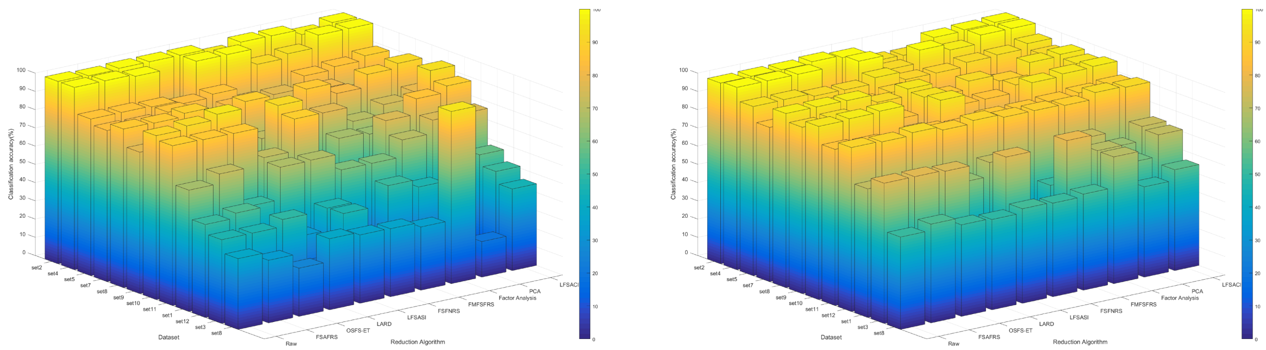
8) Principal component analysis(PCA): It aims to use the idea of dimensionality reduction to transform multiple indicators into a few comprehensive indicators.

Moreover, in order to illustrate the effectiveness of local reduction algorithm, we further compare the selection results of designed algorithm and its corresponding global algorithm GFSASI in section B. All the algorithms mentioned are run on a personal computer with Intel(R) Core(TM) i5-1135G7 CPU@2.40GH 2.42GH, and 16 GB memory. The reduction and classification algorithms are using the software MATLAB 2016b.

The twelve datasets from UCI Machine Learning Repository are adopted to conduct numerical experiments, whose detail information is shown in Table II. The the values of conditional attributes are first normalized into interval [0,1]. It is noted that the neighborhood radius is an important parameter that will influence feature selection results, thus we set δ to vary from 0.01 to 0.1 with a step of 0.01 for selecting the optimal neighborhood radius. Moreover, in order to further verify the advantages of the LFSACIE algorithm, two classical classifiers Decision Tree(DT) and K-Nearest Neighbors(KNN) based on Gini index and five neighborhoods are selected to estimate the classification accuracy and robustness of reduction algorithms, respectively. In the classification experiments, we adopt ten-folds cross validation to evaluate the different feature subset, each dataset is split to ten equal portions in each experiment, nine of them are used as the training model of the training set and the remaining one is used as the test set.

B. *Experimental Analysis*

We mainly verify the effectiveness of designed algorithms from three aspects: 1) the process of optimal neighborhood radius; 2) the comparison between local algorithm and global algorithm; and 3) the comparison between LFSACIE and other feature selection algorithms. Note that, the Raw represents the experimental results about original data.



(a) Decision Tree (b) KNN

Fig. 6. The classify performance of different reduction algorithms on two classifiers.

TABLE VI
THE FEATURE SELECTION RESULTS OF PROPOSED ALGORITHM FROM GLOBAL AND LOCAL VIEWPOINTS

Datasets	δ	GFSACIE	LFSACIE
set1	0.04	1 4 6 8	1 4 6 8
set2	0.01	1 17	1 17
set3	0.05	1 2 3 9	1 2 3 9
set4	0.01	1 3	1 3
set5	0.08	1 2 5 13	1 2 5 13
set6	0.08	8 24 38 44 50 69 82	1 13 25 39 44 53 55 59 69 82
set7	0.06	1 4 13 22 23	1 4 13 22 23
set8	0.01	1 2 4	1 2 4
set9	0.07	1 22 28 29	1 22 28 29
set10	0.06	1 17 18 19	1 17 18 19
set11	0.02	1 2 7	1 7
set12	0.07	1 2 4 6	2 3 4 6

1) The Reduction Performance of LFSACIE Under Different Neighborhood Radius: Table III records the number of selected features under different radius and Table IV records the corresponding classification performance on Decision Tree, KNN classifers. The aim of feature selection is to reduce the redundant features and further improve the classification performance, thus we select the optimal radius of different datasets according to the maximum average accuracy. Moreover, the average accuracy of different datasets is shown in Fig. 5, where the red dotted line represents the optimal radius. The optimal radius is selected for feature selection in proposed LFSACIE algorithm.

2) The Comparison Between LFSACIE and GFSACIE: Compared with global rough set, the local rough set model ignores the unnecessary computing out of target concept, which could significantly improve the compute efficiency. The consume time of LFSACIE and GFSACIE algorithm is shown in Table V. From this table, we know the consume time of LFSACIE is lower than GFSACIE in each datasets, and the time gap between two algorithms becomes bigger with the increases of data scale. Meanwhile, the selected features induced by these two algorithms are same except the sets 6, 11, 12 according to the Table VI. For the three datasets, we evaluate their feature subsets on DT and KNN classifiers, and find the average accuracy of subsets induced by LFSACIE is higher than that of GFSACIE. Therefore, we could obtain that the LFSACIE is an efficient method for feature selection compared with GFSACIE.

3) The Comparison Between LFSACIE and Other Feature Dimension Methods: There are six feature selection algorithms and two feature dimension methods are compared with the proposed algorithm LFSACIE, and the comparison between them is mainly from two aspects, the number of selected features and the corresponding classification performance. From Table VII, we find the average number of selected features for all feature selection algorithms is smaller than that of original data, thus all the algorithms can achieve dimensionality reduction. The average number of selected features of proposed algorithm is 4, ranking second out of ten compared algorithms. In order to appraise these feature subsets, we further observe the classification accuracy of them on DT and KNN classifiers, the detail comparison is shown in Fig. 6. We could find that the bar of LFASCIE is significantly higher than that of other algorithms except set 7 and set 8 on Decision Tree classifier, and the height of the bar of LFASCIE is also better than that of other algorithms on KNN classifier in most datasets.

Moreover, the more detail information is recorded in Tables VIII and IX. On Decision Tree classifier, LFSACIE achieves the highest accuracy ten times in twelve datasets, and the average accuracy is also the maximum value among all the compared algorithms and original data, thus the LFSACIE could select an excellent feature subset that can achieve a better classify performance. Similarly, the accuracy performance of LFSACIE on KNN classifier is also superior than that of other algorithms. There are eight datasets where LFSACIE behaves better than other algorithms among twelve datasets. Meanwhile, the average accuracy is 87.54 that is the highest value, and the average error is 4.53 that is the lowest value, those above classify performance shows that the LFSACIE is an excellent and robust method for feature selection.

In order to test whether there exists significant difference between different algorithms in classify performance, we adopt the Wilcoxon pairwise test to compare these experimental results. Given the test threshold is 0.1, we could find that all the test P-values are smaller than threshold according to Table X, thus we could reject the null hypothesis and consider there is

TABLE VII
THE NUMBER OF FEATURES SELECTED INDUCED BY DIFFERENT ALGORITHMS

Datasets	RAW	FSAFRS	OSFS-ET	LARD	LFSASI	FSFNRS	FMFSFRS	Factor Analysis	PCA	LFSACIE
set1	13	12	6	3	10	1	1	3	9	4
set2	18	7	6	2	9	10	17	10	8	2
set3	10	9	1	2	7	1	4	6	6	4
set4	18	17	5	2	1	2	1	10	7	2
set5	18	17	7	2	5	1	1	10	15	4
set6	101	7	5	90	17	20	100	60	1	10
set7	34	33	13	4	9	1	1	20	15	5
set8	13	12	8	2	12	1	12	7	6	3
set9	31	30	12	2	9	1	30	18	6	4
set10	23	21	6	2	6	2	22	13	5	4
set11	14	13	6	2	6	1	13	8	6	2
set12	8	7	1	2	5	1	1	4	3	4
Average	25.08	15.42	6.33	9.58	8.00	3.50	16.92	14.08	7.25	4.00

TABLE VIII
THE CLASSIFY PERFORMANCE OF DIFFERENT ALGORITHMS ON DECISION TREE CLASSIFIER

Datasets	RAW	FSAFRS	OSFS-ET	LARD	LFSASI	FSFNRS	FMFSFRS	Factor Analysis	PCA	LFSACIE
set1	87.63(6.1)	87.33(4.91)	87.15(4.12)	70.39(7.14)	87.96(4.58)	16.58(5.85)	16.58(5.85)	76.27(17.45)	84.69(4.45)	**88.13(4.9)**
set2	**100(0)**	99.48(1.09)	**100(0)**	**100(0)**	**100(0)**	96.51(2.44)	**100(0)**	99.87(0.41)	97.02(2.45)	**100(0)**
set3	50.98(16.83)	53.64(12.44)	22.27(12.03)	33.48(13.61)	45.61(15.07)	26.21(14.06)	50.76(10.8)	46.44(13.11)	36.97(14.31)	**56.89(9.82)**
set4	**100(0)**	**100(0)**	98.89(3.51)	51.11(14.77)	**100(0)**	**100(0)**	91.11(8.76)	93.33(7.77)	**100(0)**	**100(0)**
set5	85.93(4.55)	85.93(4.55)	85.37(5.05)	83.52(5.27)	85.19(5.92)	77.78(5.45)	77.78(5.45)	84.44(4.8)	85.93(4.29)	**88.33(4.28)**
set6	38.82(8.7)	34.32(4.55)	26.42(5.27)	38.98(5.25)	37.47(7.56)	37.29(6.52)	34.67(6.23)	**94.38(3.91)**	19.48(4.62)	44.72(5.98)
set7	**82.9(3.6)**	82.63(6.36)	85.76(5.36)	73.26(7.85)	81.76(5.91)	10.83(6.3)	10.83(6.3)	80.64(3.37)	74.9(7.78)	80.9(5.64)
set8	45.92(11.57)	45.92(11.57)	49.83(8.11)	20.68(6.72)	45.92(11.57)	9.39(7.35)	15.85(7.47)	10.64(8.96)	11.23(6.38)	**51.06(8.98)**
set9	87(4.97)	86.63(6.71)	88.4(7.02)	82.08(7.09)	87.7(5.9)	71.87(4.43)	87.33(6.85)	81.37(5.18)	88.22(3.89)	**90.68(3.22)**
set10	77.53(10.96)	82.58(8.08)	75.66(13.44)	68.76(11.64)	70.63(9.79)	68.74(11)	77.53(10.96)	69.34(9.93)	74.97(8.71)	**78.5(8.95)**
set11	89.87(7.56)	91.63(8.77)	94.41(6.42)	61.18(12.46)	92.09(6.06)	43.24(14.28)	66.86(8.41)	65.62(10.1)	65.75(5.96)	**93.27(6.82)**
set12	66.55(20.33)	71.82(16.09)	50.64(16.25)	68.36(14.56)	68.64(20.66)	59.73(15.16)	59.73(15.16)	69(10.34)	**76.27(17.45)**	**76.27(11.89)**
Average	76.09(7.93)	76.83(7.09)	72.07(7.22)	62.65(8.86)	75.25(7.75)	51.51(7.74)	57.42(7.69)	72.61(7.94)	67.95(6.69)	**79.06(5.87)**

TABLE IX
THE CLASSIFY PERFORMANCE OF DIFFERENT ALGORITHMS ON KNN CLASSIFIER

Datasets	RAW	FSAFRS	OSFS-ET	LARD	LFSASI	FSFNRS	FMFSFRS	Factor Analysis	PCA	LFSACIE
set1	89.91(6.04)	89.42(3.96)	90.89(3.86)	88.78(4.7)	90.4(3.12)	88.61(3.08)	88.61(3.08)	87.91(12.1)	**91.36(4.04)**	90.72(3.7)
set2	99.09(0.88)	99.22(1.09)	99.22(1.09)	100(0)	98.7(0.86)	93.01(1.75)	92.75(1.52)	98.96(1.71)	99.61(0.62)	**100(0)**
set3	70.53(9.53)	46.29(14.92)	63.86(11.89)	63.03(18.18)	70.61(9.86)	38.33(15.3)	50.91(15.36)	72.88(12.21)	64.55(13.32)	**72.5(11.61)**
set4	**100(0)**	**100(0)**	**100(0)**	64.72(8.26)	**100(0)**	91.94(12.18)	98.89(3.51)	95.42(7.91)	**100(0)**	**100(0)**
set5	92.78(4.57)	92.96(4.57)	91.3(3.12)	91.11(4.37)	94.26(2.22)	89.63(2.79)	89.63(2.79)	91.85(4.02)	92.41(3.85)	**95(2.32)**
set6	50.66(4.3)	54.14(4.61)	50.18(6.93)	49.52(4.82)	51.83(4.29)	51.3(4.6)	52.65(4.14)	**69.48(4.23)**	49.49(7.95)	55.28(5.8)
set7	84.6(6.23)	84.63(4.65)	89.48(6.52)	86.9(4.85)	89.47(5.02)	41.81(15.71)	41.81(15.71)	83.21(4.8)	81.46(5.64)	**92.05(4.32)**
set8	76.9(6.4)	76.9(5.55)	76.29(9.13)	43.17(7.18)	76.9(5.55)	38.27(5.39)	**78.74(4.89)**	69.62(8.75)	59.27(5.37)	71.15(8.17)
set9	95.48(2.36)	95.49(2.03)	95.37(1.7)	90.86(4.45)	94.56(3.14)	87.34(2.99)	91.74(2.99)	88.22(4.97)	94.72(2.87)	**95.6(1.5)**
set10	91.24(6.43)	92.24(5.07)	87.13(6.18)	84.53(9.56)	93.37(6.83)	85.08(10.24)	85.11(10.74)	89.32(5.46)	85.68(6.6)	**93.84(3.3)**
set11	94.97(4.87)	95.56(5.74)	**96.11(4.57)**	80.88(7.55)	94.93(6.19)	66.83(11.33)	68.5(15.06)	93.82(6.18)	95.05(3.78)	95.56(5.11)
set12	85.64(12.68)	85(10.96)	79.27(9.88)	85.09(10.66)	84(10.91)	86(14.39)	86(14.39)	86.73(12.72)	87.91(12.1)	**88.82(8.53)**
Average	85.98(5.36)	84.32(5.26)	84.93(5.41)	77.38(7.05)	86.59(4.83)	71.51(8.31)	77.11(7.85)	85.62(7.09)	83.46(5.51)	**87.54(4.53)**

TABLE X
THE WILCOXON TEST RESULTS OF LFSACIE AND OTHER COMPARED ALGORITHMS

Datasets \ Algorithms	RAW	FSAFRS	OSFS-ET	LARD	LFSASI	FSFNRS	FMFSFRS	Factor Analysis	PCA
Decision Tree	0.014	0.083	0.054	0.001	0.010	0.001	0.001	0.052	0.002
KNN	0.014	0.083	0.054	0.001	0.010	0.001	0.001	0.052	0.002

a significant difference between LFSACIE and other feature selection algorithms.

C. Experimental Discussion

The above experimental analysis show that the local feature selection method based on composite information entropy in fuzzy decision dataset is excellent in select efficiency and classification performance.

1) In algorithm LFSACIE, the feature evaluation index CIE first considers the deterministic and possible relative information by the upper and lower approximation ratio rather than just the specific information of one of them, which allows us to choose relatively few attributes to describe the same uncertain information as the original data.

2) In addition, the composite information entropy CIE further considers the distribution of decision attribute, which makes the selected feature more reasonable. Therefore, the classification performance and robustness of the selected feature subset based on CIE are better than those of other algorithms.

3) Finally, this algorithm LFSACIE adopts a forward heuristic selection mechanism, which greatly improves the efficiency of reduction. Besides, the composite information entropy CIE only considers the target information and ignores the other redundant information in universe, it reduces some unnecessary computation compared with global ones.

It is noted that the consume time of obtaining features based on CIE still needs to be improved due to the complexity of index. Meanwhile, this mechanism is built on static data, it needs to repeat all the additional and existing information when adding objects or features, which is time-consuming for selecting the optimal feature subset, especially in large-scale dataset. Generally, the updated dataset is related to the original ones. If we can calculate the updated data information based on their relationship, it will greatly improve the computational efficiency. Therefore, the feature selection mechanism for dynamic data also needs further to be explored.

VII. CONCLUSION

With the development of information science, the scale of data is getting bigger and a sample could be described by multiple attributes. In most practical situations, it is not a wise choice to make decisions based on all properties due to the existence of redundant attributes, thus the task of choosing important them is necessary in fuzzy decision dataset. In order to overcome the limitation of feature selection based on lower approximation in classical rough set, the local feature selection method related to composite information entropy measure is proposed in this paper, meanwhile, its corresponding heuristic algorithm LFSACIE is designed. The numerical experiment results on twelve public datasets show that the LFSACIE is an accurate and robust method which could select relative fewer features to approximate target decision and achieve better classification performances. Moreover, the LFSACIE greatly reduces consume time compared with GFSACIE according to the analysis of time complexity and experimental results. In conclusion, the LFSACIE is a relatively better method for feature selection. Based on the research results in this paper, the compute process of feature indicator and the selection mechanism for dynamic data can be further explored.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or persona relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions.

REFERENCES

[1] O. Aremu *et al.*, "A relative entropy based feature selection framework for asset data in predictive maintenance," *Comput. Ind. Eng.*, vol. 145, 2020, Art. no. 106536, doi: 10.1016/j.cie.2020.106536.

[2] B. Barman and S. Patra, "A novel technique to detect a suboptimal threshold of neighborhood rough sets for hyperspectral band selection," *Soft Comput.*, vol. 23, no. 24, pp. 13709–13719, 2019.

[3] P. P. Bonissone, A. Varma, K. S. Aggour, and F. Xue, "Design of local fuzzy models using evolutionary algorithms," *Comput. Statist. Data Anal.*, vol. 51, no. 1, pp. 398–416, 2006.

[4] H. Chen *et al.*, "Parallel attribute reduction in dominance-based neighborhood rough set," *Inf. Sci.*, vol. 373, pp. 351–368, 2016.

[5] Y. Guo, E. C. C. Tsang, W. Xu, and D. Chen, "Local logical disjunction double-quantitative rough sets," *Inf. Sci.*, vol. 500, pp. 87–112, 2019.

[6] Y. Huang *et al.*, "Matrix representation of the conditional entropy for incremental feature selection on multi-source data," *Inf. Sci.*, vol. 591, pp. 263–286, 2022.

[7] B. Huang, H. Li, and D. Wei, "Dominance-based rough set model in intuitionistic fuzzy information systems," *Knowl.-Based Syst.*, vol. 28, pp. 115–123, 2012.

[8] B. Huang, D. Wei, H. Li, and Y. Zhuang, "Using a rough set model to extract rules in dominance-based interval-valued intuitionistic fuzzy information systems," *Inf. Sci.*, vol. 221, pp. 215–229, 2013.

[9] Z. Jiang *et al.*, "Accelerator for supervised neighborhood based attribute reduction," *Int. J. Approx. Reason.*, vol. 119, pp. 122–150, 2020.

[10] W. Li *et al.*, "Distance-based double-quantitative rough fuzzy sets with logic operations," *Int. J. Approx. Reason.*, vol. 101, pp. 206–233, 2018.

[11] W. Li *et al.*, "Updating approximations with dynamic objects based on local multigranulation rough sets in ordered information systems," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 1821–1855, 2022, doi: 10.1007/s10462-021-10053-9.

[12] W. Li and W. Xu, "Double-quantitative decision-theoretic rough set," *Inf. Sci.*, vol. 316, pp. 54–67, 2015.

[13] Y. Peng and Q. Zhang, "Uncertainty measurement for set-valued data and its application in feature selection," *Int. J. Fuzzy Syst.*, vol. 24, no. 3, pp. 1735–1756, 2022, doi: 10.1007/s40815-021-01230-7.

[14] Y. Qian, C. Dang, J. Liang, and D. Tang, "Set-valued ordered information systems," *Inf. Sci.*, vol. 179, no. 16, pp. 2809–2832, 2009.

[15] Y. Qian *et al.*, "Local rough set: A solution to rough data analysis in big data," *Int. J. Approx. Reasoning*, vol. 97, pp. 38–63, 2018.

[16] B. Sang *et al.*, "Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set," *Knowl.-Based Syst.*, vol. 227, pp. 107–223, 2021.

[17] Y. Song *et al.*, "Divergence-based cross entropy and uncertainty measures of Atanassov's intuitionistic fuzzy sets with their application in decision making," *Appl. Soft Comput.*, vol. 84, 2019, Art. no. 105703, doi: 10.1016/j.asoc.2019.105703.

[18] L. Sun *et al.*, "Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy data," *IEEE Trans. Fuzzy Syst.*, vol. 29 no. 1, pp. 19–33, Jan. 2021.

[19] C. E. Shannon, "A mathsmatical theory of communication," *Bell Labs Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[20] T. K. Sheeja and A. S. Kuriakose, "A novel feature selection method using fuzzy rough sets," *Comput. Ind.*, vol. 97, pp. 111–116, 2018.

[21] L. Sun, L. Wang, W. P. Ding, Y. H. Qian, and J. C. Xu, "Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems," *Knowl.-Based Syst.*, vol. 192, 2020, Art. no. 105373, doi: 10.1016/j.knosys.2019.105373.

[22] G. Wang, X. Ma, and H. Yu, "Monotonic uncertainty measures for attribute reduction in probabilistic rough set model," *Int. J. Approx. Reasoning*, vol. 59, pp. 41–67, 2015.

[23] C. Wang *et al.*, "Fuzzy rough set-based attribute reduction using distance measures," *Knowl.-Based Syst.*, vol. 164, pp. 205–212, 2019.

[24] C. Wang *et al.*, "Feature selection based on neighborhood self-information," *IEEE Trans. Cyber.*, vol. 50, no. 9, pp. 4031–4042, Sep. 2020.

[25] C. Wang *et al.*, "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.

[26] Q. Wang, Y. Qian, X. Liang, Q. Guo, and J. Liang, "Local neighborhood rough set," *Knowl.-Based Syst.*, vol. 153, pp. 53–64, 2018.

[27] C. Wang *et al.*, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowl.-Based Syst.*, vol. 111, pp. 173–179, 2016.

[28] X. Yang *et al.*, "Neighborhood rough sets with distance metric learning for feature selection," *Knowl.-Based Syst.*, vol. 224, 2021, Art. no. 107076, doi: 10.1016/j.knosys.2021.107076.

[29] L. Yang *et al.*, "Dynamic fuzzy neighborhood rough set approach for interval-valued information systems with fuzzy decision," *Appl. Soft Comput.*, vol. 111, 2021, Art. no. 107679, doi: 10.1016/j.asoc.2021.107679.

[30] Y. Yao and X. Zhang, "Class-specific attribute reducts in rough set theory," *Inf. Sci.*, vol. 418, pp. 601–618, 2016.

[31] K. Yuan *et al.*, "An incremental learning mechanism for object classification based on progressive fuzzy three-way concept," *Inf. Sci.*, vol. 584, pp. 127–147, 2022.

[32] X. Zhang, Y. Fan, and J. Yang, "Feature selection based on fuzzy-neighborhood relative decision entropy," *Pattern Recognit. Lett.*, vol. 146, pp. 100–107, 2021.

[33] X. Zhang *et al.*, "Active incremental feature selection using a fuzzy-rough-set-based information entropy," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 901–915, May 2020.

[34] J. Zhang, X. Zhang, W. Xu, and Y. Wu, "Local multigranulation decision-theoretic rough set in ordered information systems," *Soft Comput.*, vol. 23, no. 4, pp. 13247–13261, 2019.

[35] P. Zhou *et al.*, "Online early terminated streaming feature selection based on rough set theory," *Appl. Soft Comput.*, vol. 113, 2021, Art. no. 107993, doi: 10.1016/j.asoc.2021.107993.

Weihua Xu received the Ph.D. degree from the School of Sciences, Xi'an Jiaotong University, Xi'an, China, in 2007, and the M.Sc. degree from the School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in 2004. He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has authored or coauthored three monographs and more than 120 articles in international journals. His current research interests include granular computing, approximate reasoning, fuzzy sets, rough sets, concept lattices, and cognitive computing. He is on the editorial boards of several international journals.

Kehua Yuan received the B.Sc. degree from the School of Mathematics and Statistics, Ludong University, Yantai, China, in 2019. She is currently working toward the M.Sc. degree from the College of Artificial Intelligence, Southwest University, Chongqing, China. Her current research interests include rough sets, attribute reduction, and granular computing.

Wentao Li received the M.Sc. degree from the School of Mathematics and Statistics, Chongqing University of Technology, Chongqing, China, in 2015, and the Ph.D. degree from the Department of Mathematics, Harbin Institute of Technology, Harbin, China, in 2019. He is currently a Lecturer with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has authored or coauthored more than 20 articles in international journals. His current research interests include fuzzy sets, rough sets, and granular computing. He is on the Editorial Boards of two international journals.

Weiping Ding received the Ph.D. degree in computation application from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. He is currently a Full Professor with the School of Information Science and Technology, Nantong University, Nantong, China. He was a Visiting Scholar with the University of Lethbridge, Lethbridge, AB, Canada, in 2011. From 2014 to 2015, he was a Postdoctoral Researcher with Brain Research Center, National Chiao Tung University, Hsinchu, Taiwan. In 2016, he was a Visiting Scholar with the National University of Singapore, Singapore. From 2017 to 2018, he was a Visiting Professor with the University of Technology Sydney, Ultimo, NSW, Australia. He has authored or coauthored more than 80 research peer-reviewed journal and conference papers, including IEEE T-Fs, T-NNLS, T-CYB, T-SMCS, T-BME, T-II, T-ETCI, and T-ITS. His main research interests include involve data mining, granular computing, evolutionary computing, machine learning, and Big Data analytics. He has three ESI highly cited papers. Dr. Ding was on the Editorial Advisory Board of *Knowledge-Based Systems*, and Editorial Board of *Information Fusion* and *Applied Soft Computing*. He was an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEM, *Information Sciences*, IEEE/CAA JOURNAL OF AUTOMATICA SINICA, *Neurocomputing*, *Swarm and Evolutionary Computation*, and IEEE ACCESS AND JOURNAL OF INTELLIGENT & FUZZY SYSTEMS, and the Co-Editor-in-Chief of the *Journal of Artificial Intelligence and System*.