This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

Incremental Feature Selection Using a Conditional Entropy Based on Fuzzy Dominance Neighborhood Rough Sets

Binbin Sang, Hongmei Chen*, *Member, IEEE*, Lei Yang, Tianrui Li, *Member, IEEE*, and Weihua Xu

***Abstract*—Incremental feature selection approaches can improve the efficiency of feature selection used for dynamic datasets, which has attracted increasing research attention. Nevertheless, there is currently no work on incremental feature selection approaches for dynamic ordered data. Moreover, the monotonic classification effect of ordered data is easily affected by noise, so a robust feature evaluation metric is needed for feature selection algorithm. Motivated by these two issues, we investigate incremental feature selection approaches using a new conditional entropy with robustness for dynamic ordered data in this study. First, we propose a new rough set model, $i.e.$, fuzzy dominance neighborhood rough sets (FDNRS). Second, a conditional entropy with robustness is defined based on FDNRS model, which is used as evaluation metric for features and combined with a heuristic feature selection algorithm. Finally, two incremental feature selection algorithms are designed on the basis of the above researches. Experiments are performed on ten public datasets to evaluate the robustness of the proposed metric and the performance of the incremental algorithms. Experimental results verify that the proposed metric is robust and our incremental algorithms are effective and efficient for updating reducts in dynamic ordered data.**

***Index Terms*—Incremental feature selection, fuzzy dominance neighborhood rough sets, dynamic ordered data**

I. Introduction

FEATURE selection, as a common data preprocessing approach, has elicited widespread attention in data mining [1]–[5]. This approach aims to remove redundant features from complex data and achieve the goals of reducing dimensionality, avoiding overfitting, thereby saving the time and space cost of calculation. With the development of the information age, feature selection methods have been continuously improved and innovated as the complexity and diversity of data structures increase. In real-life applications, datasets usually exhibit dynamic characteristics over time-evolving, $i.e.$, dynamic datasets. This promotes the development of incremental approaches for feature selection [6]–[10]. Incremental mechanisms of updating feature subset are widely studied, since they can effectively and efficiently fulfil feature selection tasks for dynamic datasets. However, the existing incremental approaches do not consider the monotonous ordered relation of samples in dynamic datasets. Motivated by this issue, this study focuses on investigating incremental feature selection approaches for dynamic ordered datasets.

Rough set theory (RST) proposed by Pawlak serves as an effective mathematical tool for dealing with inconsistent and uncertain information, which is a completely data-driven approach and does not require any prior knowledge of data [11]. RST is an important theoretical basis for feature selection [12]–[15]. However, in ordinal classification tasks, RST ignores the dominance principle, which requires that objects with better descriptions should not get worse labels. To offset this deficiency, Greco et al. proposed dominance-based rough set approach (DRSA) [16], which has been widely used in classification and decision-making for datasets with preference-ordered relation [17].

However, DRSA model is not robust because the knowledge granules which are constructed by considering rigorous preference-ordered relation between objects are easily affected by noise. These knowledge granules are more sensitive to noise when processing numerical data with ordered relation. In this case, the little fluctuations brought by different uncertain elements in measure and record may easily change the relations between objects, which may change the information granules and eventually obstruct users to make a correct decision. Thus, the monotonic classification and decision-making effects of ordered data are easily affected by noise. Therefore, investigating extended DRSA models to improve the robustness of DRSA is an important research work. Dominance-based neighborhood rough set (DNRS) [18] and fuzzy dominance rough set (FDRS) [19] are two important extended DRSA models. In DNRS, a dominance relation with distance was given, which qualitatively and quantitatively defines the preference-ordered relation between objects in ordered data. But the change of the consistency degree of objects ranking in ordered data cannot be effectively reflected. Because the neighborhood dominance relation followed by objects in DNRS model is a boolean relation. Hence, the degree of preference between objects cannot obtain. FDRS model considers the preference degree between objects, but the effect of noise does not be considered. Therefore, it is very meaningful to integrate the two models

This work was supported by the National Natural Science Foundation of China (Nos. 61976182, 61572406, 61602327, 61876157, and 61976245), Key Program for International S&T Cooperation of Sichuan Province (No. 2019YFH0097), and Applied Basic Research Programs of Science and Technology Department of Sichuan Province (No. 2019YJ0084). (Corresponding author: Hongmei Chen)

B.B. Sang, H.M. Chen*, and T.R. Li are with the School of Information Science and Technology, Institute of Artificial Intelligence and National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: sangbinbin@my.swjtu.edu.cn; hmchen@swjtu.edu.cn; trli@swjtu.edu.cn).

L. Yang is with the School of Mathematics, Southwest Jiaotong University, Chengdu 611756, China (e-mail: cqyanglei@yeah.net)

W.H. Xu is with the School of Artificial Intelligence, Southwest University, Chongqing 400715, China (e-mail: datongxuweihua@126.com)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

to process ordered data with noise. Inspired by this, we propose the FDNRS model, which comprehensively considers the preference degree between objects and the negative effect of noise.

Uncertainty metrics play a key role in feature selection approaches to evaluate the importance of features and quantify the inconsistency in data. Information entropy proposed by Shannon [20] has been widely concerned. Researches on information entropy have been studied extensively in different domains. For ordered data, Hu et al. proposed rank conditional entropy and fuzzy rank conditional entropy [21], and then they were applied to feature selection [22] and decision trees [23] for monotonic classification tasks. These two metrics are used to evaluate the consistency degree of the ordering of samples under features and decisions in an ordered data. However, these two metrics are sensitive to noise, which will reduces the performance of feature selection algorithms. Therefore, it is necessary to introduce a robust metric. To solve this issue, this study introduces a fuzzy dominance neighborhood conditional entropy (FDNCE) based on the proposed FDNRS model.

Feature selection methods based on DRSA have been extensively studied in the past decades, and they are used to deal with static ordered dataset [24]–[27]. Although these methods can effectively remove redundant features from ordered data, they ignore the dynamic property that the ordered data usually evolve over time in real-life applications. For dynamic ordered datasets, employing these existing approaches to compute reducts are very time-consuming, since they need to recalculate knowledge from scratch when the dataset changes slightly. This defect increases the cost of calculation space and time. Accordingly, an effective and efficient feature selection method is urgently requested to process dynamic ordered datasets.

Incremental learning is an efficient approach, which can quickly acquire new knowledge from dynamic datasets by utilizing previous knowledge [28]–[31]. In the past decade, scholars have proposed numerous incremental learning algorithms for feature selection, which mainly focus on the variations of object sets, feature sets, and feature values in a dynamic information table.

For the variation of object sets, Zhang et al. developed a fuzzy information entropy based incremental feature selection approach by using an active object screening strategy [32]. Giang et al. proposed some new incremental attribute reduction methods using the hybrid filter wrapper with fuzzy partition distance [33]. Yang et al. presented incremental updating feature subset approaches with an active object screening strategy [34], [35] and an incremental feature selection method for dynamic heterogeneous data [36]. Shu et al. introduced an incremental feature selection algorithm for dynamic hybrid data [37]. For fused decision tables, Ye et al. designed an incremental updating feature subset method via using the pseudo value of discernibility matrix [38]. Das et al. proposed a group incremental feature selection algorithm by using genetic algorithm [39]. Sang et al. designed DNRS model based heterogeneous feature selection methods with incremental mechanism for dynamic ordered data [40]. Based on fuzzy rough set theory, Ni et al. developed an incremental feature selection method that considers a key instance set containing representative instances [41].

For the variation of feature sets, Chen et al. proposed a discernible relations based incremental attribute reduction method while adding attributes [42]. Wang et al. designed an incremental feature selection algorithm via updating information entropy when the feature set vary [43]. For covering information tables, Lang et al. proposed dynamic updating feature subset methods via using related families [44]. Based on fuzzy rough set, Zeng et al. studied an incremental updating reducts algorithm on heterogeneous information table [45].

For the variation of feature values, Wei et al. introduced an incremental updating feature subset algorithm via using discernibility matrix [46], and then they developed an accelerating incremental algorithm via using a kind of compressing decision table [47]. Cai et al. studied dynamic updating reducts algorithms for a covering information table with time-evolving feature values [48]. Furthermore, Dong and Chen designed a novel RST-based incremental attribute reduction algorithm for decision table with simultaneously increasing samples and attributes [49].

It should be found that the aforementioned incremental feature selection algorithms rarely consider dynamic datasets with a preference order relation. Thence, the existing incremental feature selection algorithms are not suitable for dynamic ordered datasets, which motivates this study. Based on the above discussions, this work proposes incremental feature selection approaches for dynamic ordered datasets with time-evolving objects under the framework of FDNRS model. Different from [40], this paper improves the DNRS model and proposes a robust rough set model (*i.e.*, FDNRS model). Then, a robust feature evaluation metric and corresponding incremental feature selection algorithms are proposed based on the FDNRS model. The main difference between the literature [41] and this study is that the former considers the similarity relation between samples, while this study considers the preference relation between samples, that is, this study deals with datasets with preference relation. The major contributions of this study are as follows.

- We propose a new rough set model FDNRS, which combines the advantages of DNRS and FDRS. The proposed model is fault-tolerant for ordered data with noise, it can not only describe the relation between objects qualitatively and quantitatively, but also effectively quantify the degree of preference between objects. The polices of this model are consistent with human reasoning and meet the requirements of practical application.
- In FDNRS model framework, we define a robust uncertainty metric FDNCE, which is used to measure the degree of ranking consistency of objects in an ordered data. The property of FDNCE is presented and proved. Then, feature selection method based on FDNCE and heuristic feature selection strategy is given.
- Based on the above researches, we propose two incremental feature selection algorithms, which are used to accelerate the completion of feature selection tasks in dynamic ordered datasets.
- Comparison experiments are performed on public datasets. The robustness of the proposed metric FDNCE, and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

effectiveness and efficiency of the proposed incremental algorithms are verified by the experimental results.

The rest of this paper is organized as follows. Section II reviews preliminary knowledge on DNRS. In Section III, we construct FDNRS model. Section VI proposes FDNCE and a FDNCE-based heuristic feature selection algorithm. In Section V, two incremental approaches for feature selection are introduced. The results of our experiments are reported in Section VI. Finally, Section VII summarizes the study and outlines the further work.

II. Preliminaries

In this section, some basic concepts are introduced, which can be found in literatures [11], [17] and [18].

A. Dominance-based neighborhood rough set

1) The ordered decision system:

Definition 1: [11] Let $S = \langle U, A \cup \{d\}, V\rangle$ be a decision system, where $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty finite set of objects; A is a nonempty finite set of conditional attributes, d is a decision attribute; $V = \bigcup V_{a_k}$ $(a_k \in A \cup \{d\})$, $V_{a_k} = \{v(x_i, a_k) | \forall x_i \in U\}$, $v(x_i, a_k)$ is the value of x_i under attribute a_k, also denoted by v_{ik}.

Definition 2: [17] Let $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$ be an ordered decision system (ODS), for any $a_k \in A$, V_{a_k} is completely pre-ordered by the relation $\succeq_a$: $\forall x_i, x_j \in U$, $x_i \succeq_{a_k} x_j \Leftrightarrow v(x_i, a_k) \geq v(x_j, a_k)$ (*i.e.* an increasing preference) or $x_i \succeq_{a_k} x_j \Leftrightarrow v(x_i, a_k) \leq v(x_j, a_k)$ (*i.e.* a decreasing preference).

In real-world applications, decision makers usually know the order of criterion values according to their domain or prior knowledge. For simplicity and without any loss of generality, the following we only consider criteria with increasing preferences.

2) Neighborhood dominance relation and knowledge granules in ODS:

Definition 3: [18] Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, $\forall B \subseteq A$, the neighborhood dominance relation $N^{\preceq}_{B_\delta}$ on B is defined as

$$N^{\preceq}_{B_\delta} = \{(x_i, x_j) \in U \times U | d_B(x_i, x_j) \geq \delta \wedge v(x_i, a_k) \leq v(x_j, a_k), \forall a_k \in B\}, \tag{1}$$

where $d_B(x_i, x_j) = \min_{a_k \in B} |v(x_i, a_k) - v(x_j, a_k)|$ is the distance between x_i and x_j under B, $\delta \in (0, 1]$ is neighborhood radius. Moreover, d is a classification attribute, the dominance relation on d is denoted as $D^{\preceq}_d = \{(x_i, x_j) \in U \times U | v(x_i, d) \leq v(x_j, d)\}$.

Definition 4: [18] Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, $\forall B \subseteq A$, the neighborhood dominating and neighborhood dominated sets of $x_i \in U$ in term of B are defined as

$$N^{+}_{B_\delta}(x_i) = \{x_j \in U | x_i N^{\preceq}_{B_\delta} x_j\}; \tag{2}$$

$$N^{-}_{B_\delta}(x_i) = \{x_j \in U | x_j N^{\preceq}_{B_\delta} x_i\}, \tag{3}$$

which are called knowledge granules induced by $N^{\preceq}_{B_\delta}$.

In ODS, d is a classification attribute, $U/d = \{Cl_t | t \in \{1, \ldots, T\}\} (T \leq |U|)$, where for each Cl_t be an equivalence class, and $Cl_T \succ \cdots \succ Cl_t \succ \cdots \succ Cl_1$. The upward and downward unions in DNRS are expressed as $Cl_t^{\succeq} = \bigcup Cl_{t'} (t' \geq t)$ and $Cl_t^{\preceq} = \bigcup Cl_{t'} (t' \leq t)$, where $t, t' \in \{1, \ldots, T\}$.

3) Approximations in DNRS:

Definition 5: [18] Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, $\forall B \subseteq A$ and $t \in \{1, \ldots, T\}$, the lower and upper approximations of the upward union $Cl_t^{\succeq}$ are defined as

$$\underline{N^{\preceq}_{B_\delta}}(Cl_t^{\succeq}) = \{x \in U | N^{+}_{B_\delta}(x) \subseteq Cl_t^{\succeq}\}; \tag{4}$$

$$\overline{N^{\preceq}_{B_\delta}}(Cl_t^{\succeq}) = \{x \in U | N^{+}_{B_\delta}(x) \cap Cl_t^{\succeq} \neq \emptyset\}. \tag{5}$$

Similarly, the approximates of the downward union $Cl_t^{\preceq}$ are defined as

$$\underline{N^{\preceq}_{B_\delta}}(Cl_t^{\preceq}) = \{x \in U | N^{-}_{B_\delta}(x) \subseteq Cl_t^{\preceq}\}; \tag{6}$$

$$\overline{N^{\preceq}_{B_\delta}}(Cl_t^{\preceq}) = \{x \in U | N^{-}_{B_\delta}(x) \cap Cl_t^{\preceq} \neq \emptyset\}. \tag{7}$$

From Definition 5, the lower approximation indicates that the ranking of objects in $\underline{N^{\preceq}_{B_\delta}}(Cl_t^{\succeq})$ ($\underline{N^{\preceq}_{B_\delta}}(Cl_t^{\preceq})$) is consistent with that of in $Cl_t^{\succeq}$ ($Cl_t^{\preceq}$), the upper approximation indicates that the ranking of objects in $\overline{N^{\preceq}_{B_\delta}}(Cl_t^{\succeq})$ ($\overline{N^{\preceq}_{B_\delta}}(Cl_t^{\preceq})$) is not necessarily consistent with that of in $Cl_t^{\succeq}$ ($Cl_t^{\preceq}$).

B. Ranking problems exist in DNRS

In DRSA, the dependency reflects the consistency degree of the ranking of objects in terms of conditional attributes and decision attribute. In [18], although the DNRS model was proposed, but the corresponding dependency did not given. The following, we propose DNRS-based dependencies.

Definition 6: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, $\forall B \subseteq A$, the DNRS-based dependency of $Cl^{\succeq}$ with regard to P is defined as

$$\gamma_{B_\delta}(Cl^{\succeq}) = \frac{\sum_{t=1}^{|T|} |\underline{N^{\preceq}_{B_\delta}}(Cl_t^{\succeq})|}{\sum_{t=1}^{|T|} |Cl_t^{\succeq}|}, \tag{8}$$

where $|*|$ represents the cardinality of set $*$. Similarly, we can also define $\gamma_{B_\delta}(Cl^{\preceq})$.

However, we found that the DNRS-based dependencies cannot effectively reflect the changes in the consistency degree of the objects ranking in ODS. Here, we give an example to show this defect.

Example 1: Table I is a part of academic transcripts, where a is a conditional attribute and it represents a course, d is a decision attribute and it represents the students comprehensive level ($C \prec B \prec A$), and $x_1, x_2, \ldots,$ and x_{10} represent ten students.

TABLE I
A part of academic transcript

U	a	d	U	a	d
x_1	0.28	C	x_6	0.55	B
x_2	0.25	C	x_7	0.78	B
x_3	0.40	C	x_8	0.75	A
x_4	0.48	B	x_9	0.83	A
x_5	0.42	B	x_{10}	0.85	A

To more intuitively reflect the inconsistency of the ranking of objects with respect to a and d, we map these objects into

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

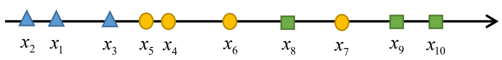


Fig. 1. The student's score ranking under course a

an axis, *i.e.*, Fig. 1, where △, ◯, and □ stand for objects coming from classes C, B, and A, respectively.

From Fig. 1, it is easy to find that the ranking of objects under a and d is inconsistent, because x_7 is assigned a relatively low level. The consistency degree of Table I can be calculated by Eq. (8) as $\gamma_{a_\delta}(Cl^{\succeq}) = 0.73$ and $\gamma_{a_\delta}(Cl^{\preceq}) = 0.83$, where $\delta = 0.1$. Suppose we respectively change the scores of objects x_3 and x_7 under a from 0.4 to 0.5 and 0.78 to 0.84, and the ranking of the revised objects is shown in Fig. 2. By comparing

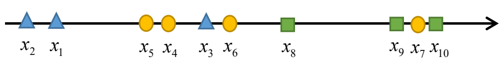


Fig. 2. The revised student's score ranking under course a

Fig. 1 and Fig. 2, we find that the degree of inconsistency in the ranking of objects becomes greater. Thence, intuitively, the DNRS-based dependencies should become smaller in this case. However, we calculated the DNRS-based dependencies of the revised version as $\gamma_{a_\delta}(Cl^{\succeq}) = 0.73$ and $\gamma_{a_\delta}(Cl^{\preceq}) = 0.83$, which are the same as the previous results. Such a result is obviously inconsistent with the logic of human reasoning.

The above analysis shows that DNRS model can not effectively reflect the change in the consistency degree of the objects ranking in an ODS. The reason lie in that the neighborhood dominance relation is a boolean relation which cannot reflect the degree of preference between objects quantitatively. The fuzzy set theory can quantify the degree of uncertainty of the concept, which meets the requirements of practical application. As pointed out by Zadeh [50], in human reasoning and concept formation, the granules used are fuzzy rather than Boolean. Therefore, we introduce fuzzy set theory into DNRS, which is necessary and meaningful.

III. Fuzzy dominance neighborhood rough sets

DNRS model provides a formal framework for studying ordered data with noise, however it cannot quantify the degree of preference for ordered data. In this section, we propose a new model, called FDNRS model, to overcome this defect. The relevant definitions are introduced as follow.

A. The fuzzy dominance neighborhood relation and fuzzy knowledge granules in ODS

Definition 7: [19] Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall a_k \in A$, and $x_i, x_j \in U$, the fuzzy dominance relation between x_i and x_j on a_k is defined as

$$\mathcal{D}^{\prec}_{a_k}(x_i, x_j) = \frac{1}{1 + e^{-k(v(x_j,a_k) - v(x_i,a_k))}}, \quad (9)$$

where k is a positive constant, and for any $B \subseteq A$, $\mathcal{D}^{\preceq}_{B}(x_i, x_j) = \min_{a_k \in B} \mathcal{D}^{\prec}_{a_k}(x_i, x_j)$.

For convenience, $\mathcal{D}^{\preceq}_{B}(x_i, x_j)$ can be simplified to $\mathcal{D}^{\prec B}_{(i,j)}$, which indicates the extent of x_j better than x_i on B. Meanwhile, a fuzzy dominance relation matrix can be formed by $\mathcal{D}^{\prec B}_{(i,j)}$, *i.e.*, $\widetilde{\mathbb{D}}^{\prec B}_{U} = [\mathcal{D}^{\prec B}_{(i,j)}]_{n \times n}$.

From Eq. (9), it is easy to find that if $v(x_j, a) > v(x_i, a)$, then $0.5 < \mathcal{D}^{\prec a}_{(i,j)} < 1$; If $v(x_j, a) = v(x_i, a)$, then $\mathcal{D}^{\prec a}_{(i,j)} = 0.5$; If $v(x_j, a) < v(x_i, a)$, then $0 < \mathcal{D}^{\prec a}_{(i,j)} < 0.5$. The fuzzy preference degree among objects calculated by using Eq. (9) are depicted in Fig. 3, where the x-coordinate denotes objects and the y-coordinate refer to the fuzzy dominance degree between other objects and the object listed in x-coordinate. It is easy to observe the distribution of fuzzy preference degree for each object.

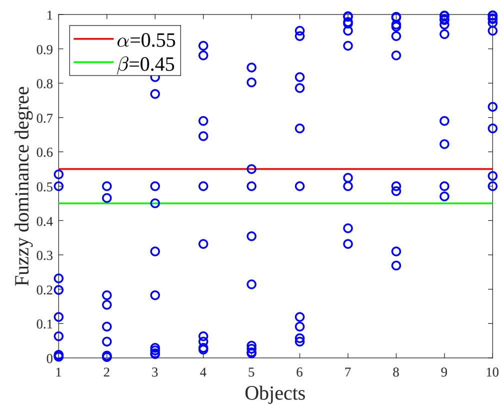


Fig. 3. The distribution of the values of fuzzy dominance relation

From Fig. 3, we can easily find that the values of fuzzy dominance relation in the area between α and β are very close to 0.5. This indicates that these objects can be regarded as no difference, because it may be caused by noise. Because in the process of collecting data, there may be a certain perturbation (*i.e.*, noise) between the real data and the collected data, which is likely to be caused by measurement tools or instruments. The knowledge granules induced by fuzzy relations may be changed by data perturbation in this case. Therefore, the definition of the fuzzy dominance neighborhood relation is proposed by adopting the strategy of neighborhood.

Definition 8: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall B \subseteq A$, and $x_i, x_j \in U$, the fuzzy dominance neighborhood relation between x_i and x_j on B is defined as

$$\mathcal{N}^{\preceq}_{B}(x_i, x_j) = \begin{cases} 0.5, & \beta \leq \mathcal{D}^{\prec B}_{(i,j)} \leq \alpha; \\ \mathcal{D}^{\prec B}_{(i,j)}, & otherwise, \end{cases} \quad (10)$$

where $\beta \in [0.4, 0.5)$, $\alpha \in (0.5, 0.6]$.

Analogously, $\mathcal{N}^{\preceq}_{B}(x_i, x_j)$ can be simplified to $\mathcal{N}^{\prec B}_{(i,j)}$, which can derive a fuzzy dominance neighborhood relation matrix, *i.e.*, $\widetilde{\mathbb{N}}^{\prec B}_{U} = [\mathcal{N}^{\prec B}_{(i,j)}]_{n \times n}$.

Definition 9: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall B \subseteq A$, the fuzzy dominating neighborhood set and fuzzy dominated neighborhood set of $x_i \in U$ in term of B are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

defined as

$$\mathcal{N}_B^{+}(x_i) = \frac{N_{(i,1)}^{\prec B}}{x_1} + \frac{N_{(i,2)}^{\prec B}}{x_2} + \cdots + \frac{N_{(i,n)}^{\prec B}}{x_n}; \quad (11)$$

$$\mathcal{N}_B^{-}(x_i) = \frac{N_{(1,i)}^{\prec B}}{x_1} + \frac{N_{(2,i)}^{\prec B}}{x_2} + \cdots + \frac{N_{(n,i)}^{\prec B}}{x_n}, \quad (12)$$

which are called fuzzy knowledge granules induced by $\mathcal{N}_{(i,j)}^{\prec B}$.

Property 1: Let $C \subseteq B \subseteq A$, then $\mathcal{N}_B^{+}(x_i) \subseteq \mathcal{N}_C^{+}(x_i)$ and $\mathcal{N}_B^{-}(x_i) \subseteq \mathcal{N}_C^{-}(x_i)$.

B. Fuzzy dominance decision in ODS

To construct FDNRS model reasonably, below we define a fuzzy dominance decision in ODS.

Definition 10: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall x_i \in U$, the fuzzy dominance decision of x_i to $Cl_t^{\succeq}$ and $Cl_t^{\preceq}(t \in \{1, \ldots, T\})$ are defined as

$$\mathcal{C}l_t^{\succeq}(x_i) = \frac{|Cl_t^{\succeq} \cap D_d^{+}(x_i)|}{|D_d^{+}(x_i)|}; \quad (13)$$

$$\mathcal{C}l_t^{\preceq}(x_i) = \frac{|Cl_t^{\preceq} \cap D_d^{-}(x_i)|}{|D_d^{-}(x_i)|}. \quad (14)$$

The $\mathcal{C}l_t^{\succeq}$ and $\mathcal{C}l_t^{\preceq}$ are two fuzzy sets, which respectively indicate the membership degree of x_i to $Cl_t^{\succeq}$ and $Cl_t^{\preceq}$.

C. Approximations in FDNRS

The upward and downward unions are then described approximately by comprehensively considering fuzzy dominance decision and fuzzy dominance neighborhood relation. The definitions of approximations are given below.

Definition 11: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall B \subseteq A$ and $t \in \{1, \ldots, T\}$, the lower and upper approximations of the upward union $Cl_t^{\succeq}$ under B are defined as

$$\underline{\mathcal{N}_B^{\prec}}(\mathcal{C}l_t^{\succeq})(x_i) = \inf_{x_j \in U} \max(1 - \mathcal{N}_B^{+}(x_i)(x_j), \mathcal{C}l_t^{\succeq}(x_j)); \quad (15)$$

$$\overline{\mathcal{N}_B^{\prec}}(\mathcal{C}l_t^{\succeq})(x_i) = \sup_{x_j \in U} \min(\mathcal{N}_B^{-}(x_i)(x_j), \mathcal{C}l_t^{\succeq}(x_j)). \quad (16)$$

Similarly, the approximates of the downward union $Cl_t^{\preceq}$ under B are defined as

$$\underline{\mathcal{N}_B^{\prec}}(\mathcal{C}l_t^{\preceq})(x_i) = \inf_{x_j \in U} \max(1 - \mathcal{N}_B^{-}(x_i)(x_j), \mathcal{C}l_t^{\preceq}(x_j)); \quad (17)$$

$$\overline{\mathcal{N}_B^{\prec}}(\mathcal{C}l_t^{\preceq})(x_i) = \sup_{x_j \in U} \min(\mathcal{N}_B^{+}(x_i)(x_j), \mathcal{C}l_t^{\preceq}(x_j)). \quad (18)$$

D. The dependency degree of $Cl^{\succeq}$ in FDNRS

Definition 12: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall B \subseteq A$, the dependency degree of $Cl^{\succeq}$ in FDNRS with regard to B is defined as

$$\widetilde{\gamma}_B(Cl^{\succeq}) = \frac{\sum_{t=1}^{|T|} \sum_{i=1}^{|U|} \underline{\mathcal{N}_B^{\prec}}(\mathcal{C}l_t^{\succeq})(x_i)}{\sum_{t=1}^{|T|} \sum_{i=1}^{|U|} \mathcal{C}l_t^{\succeq}(x_i)}. \quad (19)$$

Similarly, we can also define $\widetilde{\gamma}_B(Cl^{\preceq})$.

The following we verify whether the FDNRS based dependencies can effectively reflect the changes in the consistency of the objects ranking in ODS.

Example 2: Continuing from Example 1. The calculation results corresponding to the DNRS-based dependencies and the FDNRS-based dependencies in Figs. 1 and 2 are shown in Table II, respectively.

TABLE II
DEPENDENCIES BASED ON DNRS AND FDNRS

	Fig. 1		Fig. 2	
	γ_{a_δ}	$\widetilde{\gamma}_a$	γ_{a_δ}	$\widetilde{\gamma}_a$
$Cl^{\succeq}$	0.73	0.92	0.73	0.90 ↓
$Cl^{\preceq}$	0.83	0.95	0.83	0.93 ↓

Although the inconsistency in Fig. 2 should become larger than that of Fig. 1. From Table II, we find that there is no difference in dependencies under DNRS model. In this case, the dependencies under FDNRS model become smaller, which is more reasonable and consistent with human reasoning.

The above analysis shows that FDNRS model can effectively reflect the change in the consistency degree of the objects ranking in an ODS. Because knowledge granules in FDNRS are induced by the fuzzy neighborhood dominance relation, it can quantify the degree of preference between objects. Therefore, FDNRS model not only inherits the advantages of DNRS, but also is consistent with human reasoning and meets the requirements of practical application.

IV. CONDITIONAL ENTROPY BASED ON FDNRS AND NON-MONOTONIC FEATURE SELECTION

Information entropy is a common uncertainty measure, which performs well in feature selection tasks. In this section, we first propose a conditional entropy based on FDNRS, called FDNCE, and analyze its monotonicity. Afterwards, we define a non-monotonic reduct search strategy via using FDNCE. Finally, we introduce a heuristic feature selection algorithm with the non-monotone reduct search strategy.

A. Fuzzy dominance neighborhood conditional entropy

In [21], Hu et al. successively proposed dominance conditional entropy (DCE) and fuzzy dominance conditional entropy (FDCE) for evaluating the consistency degree of the ranking of objects under features and decisions in an ODS. Obviously, DCE follows the dominance relation, which only reflects the dominance relation between objects from the qualitative perspectives. FDCE follows the fuzzy dominance relation (as Definition 7), which reflects the dominance relation between objects from both qualitative and quantitative perspectives. However, as we mentioned earlier, the fuzzy dominance relation does not consider the effects of noise. To make up for this defect, the following we define the FDNCE in an ODS.

Definition 13: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, $\forall B \subseteq A$, the FDNCE of B relative to d is defined as

$$\mathcal{N}\mathcal{E}_{d|B}^{\prec}(U) = -\frac{1}{|U|} \sum_{i=1}^{n} \log \frac{|\mathcal{N}_B^{+}(x_i) \cap D_d^{+}(x_i)|}{|\mathcal{N}_B^{+}(x_i)|}. \quad (20)$$

Similarly, the neighborhood dominance relation based conditional entropy (NDCE) can also be defined as Eq. (20).

In Eq. (20), $\frac{|\mathcal{N}_B^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_B^+(x_i)|}$ can be regarded as a variable, which is the core part of $\mathcal{NE}_{d|B}^{\prec}(U)$. Intuitively, this variable measures the consistency degree of the objects ranking in terms of the conditional attribute set B and the decision d. It is easy to find that the value of FDNCE is inversely proportional to this variable, and $\mathcal{NE}_{d|B}^{\prec}(U)$ is non-negative. When using FDNCE to evaluate an attribute subset, it is expect that the ranking information provided by this attribute subset for the objects in ODS is the same as the decision. Therefore, the more smaller value of $\mathcal{NE}_{d|B}^{\prec}(U)$ (or the larger value of variable $\frac{|\mathcal{N}_B^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_B^+(x_i)|}$), the more meaningful of attribute subset B. Next, we prove that FDNCE is non-monotonicity.

Property 2: Let $C \subseteq B \subseteq A$, then $\mathcal{NE}_{d|C}^{\prec}(U) \leq \mathcal{NE}_{d|B}^{\prec}(U)$ or $\mathcal{NE}_{d|C}^{\prec}(U) \geq \mathcal{NE}_{d|B}^{\prec}(U)$ is indeterminate, namely, FDNCE is non-monotonic.

Proof : From Eq. 20, we have

$$\begin{aligned}\triangle &= \mathcal{NE}_{d|B}^{\prec}(U) - \mathcal{NE}_{d|C}^{\prec}(U)\\ &= \frac{1}{|U|}\sum_{i=1}^{n}(\log\frac{|\mathcal{N}_C^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_C^+(x_i)|} - \log\frac{|\mathcal{N}_B^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_B^+(x_i)|}).\end{aligned}$$

Assuming that $g_1(x_i) = \frac{|\mathcal{N}_C^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_C^+(x_i)|}$ and $g_2(x_i) = \frac{|\mathcal{N}_B^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_B^+(x_i)|}$. It can be obtained that $\triangle = \frac{1}{|U|}\sum_{i=1}^{n}(\log g_1(x_i) - \log g_2(x_i)) = \frac{1}{|U|}\sum_{i=1}^{n}\log\frac{g_1(x_i)}{g_2(x_i)}$. Since $|\mathcal{N}_C^+(x_i) \cap D_d^+(x_i)| < |\mathcal{N}_C^+(x_i)|$ and $|\mathcal{N}_B^+(x_i) \cap D_d^+(x_i)| < |\mathcal{N}_B^+(x_i)|$ hold, then $0 < g_1(x_i), g_2(x_i) < 1$ holds. Hence, $\frac{g_1(x_i)}{g_2(x_i)} > 1$ ($\frac{g_1(x_i)}{g_2(x_i)} < 1$) is uncertain. So $\triangle > 0$ ($\triangle < 0$) is indeterminate. Therefore, FDNCE is non-monotonic.

B. The evaluation of attributes in ODS

Definition 14: Given an ODS $S^{\preceq} = \langle U, A\cup\{d\}, V\rangle$, $\forall Q \subseteq A$, we say Q is a reduct of A relative to d if Q satisfies

(1) $\mathcal{NE}_{d|Q}^{\prec}(U) \leq \mathcal{NE}_{d|A}^{\prec}(U)$,

(2) $\forall a_k \in Q$, $\mathcal{NE}_{d|(Q-\{a_k\})}^{\prec}(U) > \mathcal{NE}_{d|Q}^{\prec}(U)$.

The first item guarantees that the selected attribute subset Q can provide correct objects ranking information that is not worse than that of raw attribute set A. The second item requires that no redundant attributes in the selected attribute subset Q.

According to Definition 14, we define the inner and outer significance of an attribute as follows.

Definition 15: Given an ODS $S^{\preceq} = \langle U, A\cup\{d\}, V\rangle$, $\forall B \subseteq A$ and $\forall a \in B$, the inner significance of a relative to B is defined as

$$sig_{inner}^{U}(a, B, d) = \mathcal{NE}_{d|(B-\{a\})}^{\prec}(U) - \mathcal{NE}_{d|B}^{\prec}(U). \quad (21)$$

Definition 16: Given an ODS $S^{\preceq} = \langle U, A\cup\{d\}, V\rangle$, $\forall B \subseteq A$ and $\forall a \in (C - B)$, the outer significance of a relative to B is defined as

$$sig_{outer}^{U}(a, B, d) = \mathcal{NE}_{d|B}^{\prec}(U) - \mathcal{NE}_{d|(B\cup\{a\})}^{\prec}(U). \quad (22)$$

The matrix representation of knowledge is an intuitive and effective way for processing complex data, and the calculation of the matrix can be easily implemented via using a computer. Thence, it is necessary to present a method for computing FDNCE by using relation matrices. In what follows, we define some operations on relation matrices.

Definition 17: Let $B_1, B_2 \subseteq A\cup\{d\}$, $\mathbb{R}_U^{B_1} = [r_{(i,j)}^{B_1}]_{n\times n}$ and $\mathbb{R}_U^{B_2} = [r_{(i,j)}^{B_2}]_{n\times n}$ are two relation matrices under attribute sets B_1 and B_2, respectively, then the "$\wedge$" and "$*$" operations between them are defined as

$$\mathbb{R}_U^{B_1} \wedge \mathbb{R}_U^{B_2} = [\min\{r_{(i,j)}^{B_1}, r_{(i,j)}^{B_2}\}]_{n\times n}, \quad (23)$$

$$\mathbb{R}_U^{B_1} * \mathbb{R}_U^{B_2} = [r_{(i,j)}^{B_1} \times r_{(i,j)}^{B_2}]_{n\times n}. \quad (24)$$

Definition 18: Let $B \subseteq A \cup \{d\}$, $\mathbb{R}_U^{B} = [r_{(i,j)}^{B}]_{n\times n}$ be a relation matrix, and its diagonal matrix is defined as $\widehat{\mathbb{R}_U^{B}} = [\widehat{r}_{(i,j)}^{B}]_{n\times n}$, where

$$\widehat{r}_{(i,j)}^{B} = \begin{cases} \sum_{l=1}^{n} r_{(i,l)}^{B}, & i,j \in [1,n], i = j; \\ 0, & i,j \in [1,n], i \neq j. \end{cases} \quad (25)$$

Moreover, the determinant and inverse matrix of $\widehat{\mathbb{R}_U^{B}}$ are denoted as $|\widehat{\mathbb{R}_U^{B}}| = \Pi_{i=j=1}^{n}\widehat{r}_{(i,j)}^{B}$ and $(\widehat{\mathbb{R}_U^{B}})^{-1} = [1/\widehat{r}_{(i,j)}^{B}]_{n\times n}$, respectively.

Corollary 1: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, $\forall B \subseteq A$, the formula for calculating FDNCE using matrices is expressed as

$$\mathcal{NE}_{d|B}^{\prec}(U) = -\frac{1}{|U|}\log|\widehat{\widetilde{\mathbb{N}}_U^{\prec B\cup\{d\}}} * (\widehat{\widetilde{\mathbb{N}}_U^{\prec B}})^{-1}|, \quad (26)$$

where $\widetilde{\mathbb{N}}_U^{\prec B\cup\{d\}} = \widetilde{\mathbb{N}}_U^{\prec B} \wedge \mathbb{D}_U^{\prec d} = [\mathcal{N}_{(i,j)}^{\prec B\cup\{d\}}]_{n\times n}$.

Proof : According to Eq. (26), we can get that

$$\begin{aligned}\mathcal{NE}_{d|B}^{\prec}(U) &= -\frac{1}{|U|}\log\Pi_{i=j=1}^{n}\frac{\widehat{\mathcal{N}}_{(i,j)}^{\prec B\cup\{d\}}}{\widehat{\mathcal{N}}_{(i,j)}^{\prec B}} = -\frac{1}{|U|}\log\frac{\Pi_{i=j=1}^{n}\widehat{\mathcal{N}}_{(i,j)}^{\prec B\cup\{d\}}}{\Pi_{i=j=1}^{n}\widehat{\mathcal{N}}_{(i,j)}^{\prec B}}\\ &= -\frac{1}{|U|}\log\frac{\Pi_{i=1}^{n}(\sum_{l=1}^{n}\mathcal{N}_{(i,l)}^{\prec B\cup\{d\}})}{\Pi_{i=1}^{n}(\sum_{l=1}^{n}\mathcal{N}_{(i,l)}^{\prec B})} = -\frac{1}{|U|}\log\frac{\Pi_{i=1}^{n}|\mathcal{N}_{B\cup\{d\}}^{+}(x_i)|}{\Pi_{i=1}^{n}|\mathcal{N}_{B}^{+}(x_i)|} =\\ &-\frac{1}{|U|}\log\frac{\Pi_{i=1}^{n}|\mathcal{N}_B^+(x_i)\cap D_d^+(x_i)|}{\Pi_{i=1}^{n}|\mathcal{N}_B^+(x_i)|} = -\frac{1}{|U|}\sum_{l=1}^{n}\log\frac{|\mathcal{N}_B^+(x_i)\cap D_d^+(x_i)|}{|\mathcal{N}_B^+(x_i)|}.\end{aligned}$$

From this we can conclude that the results of computing FDNCE via using Eq. (20) and Eq. (26) are equal.

C. Heuristic feature selection algorithm

In this subsection, we design a FDNCE based heuristic feature selection algorithm (FDNCE-HFS) according to Definition 14, and then analyze its time complexity.

1) FDNCE-HFS algorithm (see Algorithm 1): In algorithm FDNCE-HFS, Step 2 is to calculate FDNCE under raw attribute set A. Steps 3-9 is to add attributes with inner significance greater than zero to Red_U, and let $Q = Red_U$. Steps 10-16 is to insert the attribute with the highest outer significance from remaining attribute subset $A - Q$ into Q until Step 10 does not hold. Steps 17-22 is to delete redundant attributes from attribute subset Q. Steps 23-24 is to output the final reduct.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

Algorithm 1: FDNCE-HFS algorithm

Input: An ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, parameters α, and β.
Output: A reduct Red_U.
1: Initialize $Red_U \leftarrow \emptyset$;
2: Calculate FDNCE $\mathcal{NE}^{\preceq}_{d|A}(U)$ via using Eq. (26);
3: **for** $k = 1$ to $|A|$ **do**
4: Calculate $sig^U_{inner}(a_k, A, d)$ by Definition 15;
5: **if** $sig^U_{inner}(a_k, A, d) > 0$ **then**
6: $Red_U \leftarrow Red_U \cup \{a_k\}$;
7: **end if**
8: **end for**
9: Let $Q \leftarrow Red_U$;
10: **while** $\mathcal{NE}^{\preceq}_{d|Q}(U) > \mathcal{NE}^{\preceq}_{d|A}(U)$ **do**
11: **for** $l = 1$ to $|A - Q|$ **do**
12: Calculate $sig^U_{outer}(a_l, Q, d)$ by Definition 16;
13: **end for**
14: Select $a_0 = max\{sig^U_{outer}(a_l, Q, d), a_l \in (A - Q)\}$;
15: $Q \leftarrow Q \cup \{a_0\}$
16: **end while**
17: **for** each $a \in Q$ **do**
18: Calculate FDNCE $\mathcal{NE}^{\preceq}_{d|(Q-\{a\})}(U)$ via using Eq. (26);
19: **if** $\mathcal{NE}^{\preceq}_{d|(Q-\{a\})}(U) \leq \mathcal{NE}^{\preceq}_{d|Q}(U)$ **then**
20: $Q \leftarrow Q - \{a\}$;
21: **end if**
22: **end for**
23: $Red_U \leftarrow Q$;
24: **return** Red_U;

TABLE III
THE TIME COMPLEXITY OF ALGORITHM FDNCE-HFS

Steps	Time complexity	Steps	Time complexity
2	$O(\|A\|\|U\|^2)$	10-16	$O(\|A\|^2\|U\|^2)$
3-9	$O(\|A\|^2\|U\|^2)$	17-22	$O(\|Q\|^2\|U\|^2)$

2) Time complexity: The time complexity of the main steps in algorithm FDNCE-HFS are listed in Table III.

The heuristic feature selection method is a common feature selection strategy. Therefore, analogously, heuristic feature selection (HFS) algorithms based on DCE, NDCE, and FDCE can also be designed. In experiments, these algorithms are compared with FDNCE-HFS.

V. INCREMENTAL APPROACHES FOR FEATURE SELECTION WITH THE VARIATION OF MULTIPLE OBJECTS

For dynamic ODS with objects change, employing the FDNCE-HFS algorithm to compute a reduct is very time-consuming, especially in large data. Because this algorithm retrains the changed ODS as a new one, which needs to recalculate knowledge from scratch. To improve efficiency, this section presents two incremental algorithms for feature selection on the basis of FDNCE-HFS algorithm.

A. The updating mechanism of FDNCE when adding objects

Uncertainty metric is an important part of feature selection algorithms, and its calculation speed determines the efficiency of the algorithms. Thence, this subsection present an incremental update mechanism that is used to quickly compute the new FDNCE when adding objects to an ODS. From Eq. (26), we can easily find that the pivotal step in the process of updating FDNCE is to calculate the corresponding diagonal matrix in an incremental manner. In what follows, the principle for updating the diagonal matrix is presented.

Proposition 1: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$, adding object set $U_{ad} = \{x_{n+1}, x_{n+2}, \ldots, x_{n+n'}\}$ to $S^{\preceq}$, then the changed object set is $U' = U \cup U_{ad}$. Let $\forall B \subseteq A$, known the previous diagonal matrix is $\widehat{\widetilde{\mathbb{N}}^{\prec B}_{U}} = [\widehat{\mathcal{N}}^{\prec B}_{(i,j)}]_{n \times n}$, which is updated to $\widehat{\widetilde{\mathbb{N}}^{\prec B}_{U'}} = [\widehat{\mathcal{N}}'^{\prec B}_{(i,j)}]_{(n+n') \times (n+n')}$ after adding objects, where

$$\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \begin{cases} \widehat{\mathcal{N}}^{\prec B}_{(i,j)} + \sum\limits_{l=n+1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}, & i,j \in [1,n], i = j; \\ \sum\limits_{l=1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}, & i,j \in [n+1, n+n'], i = j; \\ 0, & i,j \in [1, n+n'], i \neq j, \end{cases} \tag{27}$$

where $\widehat{\mathcal{N}}^{\prec B}_{(i,j)}$ is known, $\sum_{l=n+1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}$ and $\sum_{l=1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}$ need to be calculated by Eq. (10).

Proof : According to Definition 18, we know that all non-diagonal elements in matrix $\widehat{\widetilde{\mathbb{N}}^{\prec B}_{U'}}$ are zero, that is, $\forall i,j \in [1, n+n']$ and $i \neq j$, $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = 0$ always holds. Then $\forall i,j \in [1,n]$ and $i = j$, we have $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \sum_{l=1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)} = \sum_{l=1}^{n} \mathcal{N}^{\prec B}_{(i,l)} + \sum_{l=n+1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)} = \widehat{\mathcal{N}}^{\prec B}_{(i,j)} + \sum_{l=n+1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}$, where $\widehat{\mathcal{N}}^{\prec B}_{(i,j)}$ is known, and $\sum_{l=n+1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}$ needs to be calculated by Eq. (10). Furthermore, $\forall i,j \in [n+1, n+n']$ and $i = j$, $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \sum_{l=1}^{n+n'} \mathcal{N}^{\prec B}_{(i,l)}$ also needs to be calculated by Eq. (10). In summary, based on the previous diagonal matrix $\widehat{\widetilde{\mathbb{N}}^{\prec B}_{U}}$, we calculate new knowledge to obtain an updated diagonal matrix $\widehat{\widetilde{\mathbb{N}}^{\prec B}_{U'}}$, where $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)}$ is denoted as Eq. (27).

Analogously, the diagonal matrix $\widehat{\widetilde{\mathbb{N}}^{\prec B \cup \{d\}}_{U'}}$ can also be updated by Proposition 1. Therefore, according to Eq. (26), we can directly compute the new FDNCE via using the updated matrices $\widehat{\widetilde{\mathbb{N}}^{\prec B}_{U'}}$ and $\widehat{\widetilde{\mathbb{N}}^{\prec B \cup \{d\}}_{U'}}$.

B. An incremental algorithm when adding objects

Based on FDNCE-HFS algorithm, this subsection introduces an incremental feature selection algorithm when adding objects (FDNCE-IFSA), and then analyze its time complexity.

1) FDNCE-IFSA algorithm (see Algorithm 2): In Algorithm 2, Step 1 is to add the object set to the original ODS. Step 2 is to update the original diagonal matrices by Proposition 1. Step 3 is to calculate the new FDNCE via using Eq. (26). Steps 4-8 is to determine whether the new FDNCE under the previous reduct Q is equal to or less than that of under the raw attribute set A, if so, then keep the previous reduct unchanged. Steps 9-14 is to construct a descending sequence for the remaining attributes, and incrementally update the selected attribute subset until Step 10 does not hold. Steps 15-20 is to remove redundant attributes from the selected attribute subset. Steps 21-22 is to output the final reduct.

Algorithm 2: FDNCE-IFSA algorithm

Input: An original ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, and its reduct Q, parameters α, β, original diagonal matrices $\widehat{\widetilde{\mathbb{N}}_U^{\prec A}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec A\cup\{d\}}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q\cup\{d\}}}$, and $U_{ad} = \{x_{n+1}, x_{n+2}, \ldots, x_{n+n'}\}$;

Output: A new reduct $Red_{U'}$ on $U \cup U_{ad}$.

1: Add object set $U' \leftarrow U \cup U_{ad}$;
2: Update the diagonal matrices $\widehat{\widetilde{\mathbb{N}}_U^{\prec A}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec A}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec A\cup\{d\}}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec A\cup\{d\}}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec Q}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q\cup\{d\}}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec Q\cup\{d\}}}$ by Proposition 1;
3: Calculate the new FDNCE $\mathcal{NE}^{\prec}_{d|A}(U')$ and $\mathcal{NE}^{\prec}_{d|Q}(U')$ via using Eq. (26);
4: **if** $\mathcal{NE}^{\prec}_{d|Q}(U') \leq \mathcal{NE}^{\prec}_{d|A}(U')$ **then**
5: turn to step 15;
6: **else**
7: turn to step 9;
8: **end if**
9: For each $a \in (A - Q)$, calculate $sig^{U'}_{outer}(a, Q, d)$ via using Eq. (22), then ranking these attributes w.r.t descending order of their outer significance, and record the results as $\{a'_1, a'_2, \ldots, a'_{|A-Q|}\}$;
10: **while** $\mathcal{NE}^{\prec}_{d|Q}(U') > \mathcal{NE}^{\prec}_{d|A}(U')$ **do**
11: **for** $h = 1$ to $|A - Q|$ **do**
12: select $Q \leftarrow Q \cup \{a'_h\}$ and calculate $\mathcal{NE}^{\prec}_{d|Q}(U')$;
13: **end for**
14: **end while**
15: **for** each $a \in Q$ **do**
16: Calculate FDNCE $\mathcal{NE}^{\prec}_{d|(Q-\{a\})}(U')$ via using Eq. (26);
17: **if** $\mathcal{NE}^{\prec}_{d|(Q-\{a\})}(U') \leq \mathcal{NE}^{\prec}_{d|Q}(U')$ **then**
18: $Q \leftarrow Q - \{a\}$;
19: **end if**
20: **end for**
21: $Red_{U'} \leftarrow Q$;
22: **return** $Red_{U'}$;

2) The time complexity of FDNCE-IFSA algorithm: The time complexity of the main steps in this algorithm are listed in Table IV.

TABLE IV
THE TIME COMPLEXITY OF ALGORITHM FDNCE-IFSA

Steps	Time complexity	Steps	Time complexity
2-3	$O(\lvert A\rvert\lvert U_{ad}\rvert\lvert U'\rvert)$	15-20	$O(\lvert Q\rvert^2\lvert U'\rvert^2)$
9-14	$O((\lvert A\rvert - \lvert Q\rvert)\lvert U'\rvert^2)$		

3) The comparison of time complexity: We list the time complexity of algorithms FDNCE-HFS and FDNCE-IFSA in Table V for intuitive comparison.

TABLE V
THE COMPARISON OF THE TIME COMPLEXITY OF ALGORITHMS FDNCE-HFS AND FDNCE-IFSA

Algorithms	Time complexity
FDNCE-HFS	$O(\lvert A\rvert\lvert U'\rvert^2 + \lvert A\rvert^2\lvert U'\rvert^2 + \lvert A\rvert^2\lvert U'\rvert^2 + \lvert Q\rvert^2\lvert U'\rvert^2)$
FDNCE-IFSA	$O(\lvert A\rvert\lvert U_{ad}\rvert\lvert U'\rvert + (\lvert A\rvert - \lvert Q\rvert)\lvert U'\rvert^2 + \lvert Q\rvert^2\lvert U'\rvert^2)$

From Table V, we can easily find that the time complexity of FDNCE-IFSA algorithm is usually much less than that of FDNCE-HFS algorithm. Because FDNCE-HFS algorithm computes a new reduct from scratch, it ignores the previously acquired knowledge. By contrast, FDNCE-IFSA algorithm uses the previous knowledge for accelerating the acquisition of a new reduct. Thence, compared with FDNCE-HFS algorithm, FDNCE-IFSA algorithm saves time cost.

C. The updating mechanism of FDNCE when deleting objects

In this subsection, we introduce an incremental update mechanism for calculating the new FDNCE when objects are deleted from an ODS.

Proposition 2: Given an ODS $S^{\preceq} = \langle U, A \cup \{d\}, V\rangle$, deleting object set $U_{de} = \{x_{q_1}, x_{q_2}, \ldots, x_{q_{n'}}\}$ from $S^{\preceq}$, then the changed object set is $U' = U - U_{de}$. Let $\forall B \subseteq A$, known the previous relation matrix $\widetilde{\mathbb{N}}_U^{\prec B} = [\mathcal{N}^{\prec B}_{(i,j)}]_{n\times n}$ and its diagonal matrix $\widehat{\widetilde{\mathbb{N}}_U^{\prec B}} = [\widehat{\mathcal{N}}^{\prec B}_{(i,j)}]_{n\times n}$, where the diagonal matrix is updated to $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B}} = [\widehat{\mathcal{N}}'^{\prec B}_{(i,j)}]_{(n-n')\times(n-n')}$ after deleting objects, where

$$\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \begin{cases} \widehat{\mathcal{N}}^{\prec B}_{(i+k-1,j+k-1)} - \sum\limits_{t=1}^{n'} \mathcal{N}^{\prec B}_{(i+k-1,q_t)}, \\ \quad i,j \in [q_{k-1} - k + 2, q_k - k + 1), i = j; \\ \widehat{\mathcal{N}}^{\prec B}_{(i+n',j+n')} - \sum\limits_{t=1}^{n'} \mathcal{N}^{\prec B}_{(i+n',q_t)}, \\ \quad i,j \in [q_{n'} - n' + 1, n - n'], i = j; \\ 0, \quad i,j \in [1, n-n'], i \neq j, \end{cases} \tag{28}$$

where $1 \leq k \leq n'$.

Proof: When the object set U_{de} is deleted, the raw object set becomes $U' = \{x_1, x_2, \ldots, x_{n-n'}\}$. In $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B}}$, the elements on the off-diagonal lines are all zero, *i.e.*, $\forall i, j \in [1, n - n']$ and $i \neq j$, $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = 0$ always holds. According to Definition 18, for elements on the diagonal, we have $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \sum\limits_{l=1}^{n} \mathcal{N}^{\prec B}_{(i,l)} - \sum\limits_{t=1}^{n'} \mathcal{N}^{\prec B}_{(i,t)} = \widehat{\mathcal{N}}^{\prec B}_{(i,j)} - \sum\limits_{t=1}^{n'} \mathcal{N}^{\prec B}_{(i,t)}$, and its position has two changes in $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B}}$. One for any $i, j \in [q_{k-1}, q_k)$ and $i = j$, the row and column coordinates of $\widehat{\mathcal{N}}^{\prec B}_{(i,j)}$ should be shifted forward by $k - 1$ positions at the same time. After that, we can get that for any $i, j \in [q_{k-1} - k + 2, q_k - k + 1)$ and $i = j$, $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \widehat{\mathcal{N}}^{\prec B}_{(i+k-1,j+k-1)} - \sum\limits_{t=1}^{n'} \mathcal{N}^{\prec B}_{(i+k-1,q_t)}$ holds. On the other hand, for any $i, j \in [q_{n'} - n' + 1, n - n']$ and $i = j$, the row and column coordinates of $\widehat{\mathcal{N}}^{\prec B}_{(i,j)}$ should be shifted forward by n' positions simultaneously. Then, we have $\widehat{\mathcal{N}}'^{\prec B}_{(i,j)} = \widehat{\mathcal{N}}^{\prec B}_{(i+n',j+n')} - \sum\limits_{t=1}^{n'} \mathcal{N}^{\prec B}_{(i+n',q_t)}$ holds. To sum up, based on the previous relation matrix $\widetilde{\mathbb{N}}_U^{\prec B}$ and its diagonal matrix $\widehat{\widetilde{\mathbb{N}}_U^{\prec B}}$, we delete the corresponding knowledge to obtain an updated diagonal matrix $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B}}$.

Analogously, the diagonal matrix $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B\cup\{d\}}}$ can also be updated by Proposition 2. Hence, according to Eq. (26), we can directly compute the new FDNCE via using the updated matrices $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B}}$ and $\widehat{\widetilde{\mathbb{N}}_{U'}^{\prec B\cup\{d\}}}$.

D. An incremental algorithm when deleting objects

Based on FDNCE-HFS algorithm, this subsection design an incremental feature selection algorithm when deleting objects (FDNCE-IFSD), and then analyze its time complexity.

Algorithm 3: FDNCE-IFSD algorithm

Input: An original $S^{\preceq} = \langle U, A \cup \{d\}, V \rangle$ and its reduct Q, parameters α, β, original relation matrices $\widetilde{\mathbb{N}}_U^{\prec A}$, $\widetilde{\mathbb{N}}_U^{\prec A\cup\{d\}}$, $\widetilde{\mathbb{N}}_U^{\prec Q}$, $\widetilde{\mathbb{N}}_U^{\prec Q\cup\{d\}}$, and their diagonal matrices $\widehat{\widetilde{\mathbb{N}}_U^{\prec A}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec A\cup\{d\}}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q\cup\{d\}}}$, and $U_{de} = \{x_{q_1}, x_{q_2}, \ldots, x_{q_{n'}}\}$;
Output: A new reduct $Red_{U'}$ on $U - U_{de}$.
1: Delete object set $U' \leftarrow U - U_{de}$;
2: Update the diagonal matrices $\widehat{\widetilde{\mathbb{N}}_U^{\prec A}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec A}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec A\cup\{d\}}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec A\cup\{d\}}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec Q}}$, $\widehat{\widetilde{\mathbb{N}}_U^{\prec Q\cup\{d\}}} \rightarrow \widehat{\widetilde{\mathbb{N}}_{U'}^{\prec Q\cup\{d\}}}$ by Proposition 2;
3: Calculate the new FDNCE $\mathcal{NE}_{d|A}^{\preceq}(U')$ and $\mathcal{NE}_{d|Q}^{\preceq}(U')$ via using Eq. (26);
4: **if** $\mathcal{NE}_{d|Q}^{\preceq}(U') \leq \mathcal{NE}_{d|A}^{\preceq}(U')$ **then**
5: turn to step 15;
6: **else**
7: turn to step 9;
8: **end if**
9: For each $a \in (A - Q)$, calculate $sig_{outer}^{U'}(a, Q, d)$ via using Eq. (22), then construct a descending sequence of attributes, and record the results as $\{a'_1, a'_2, \ldots, a'_{|A-Q|}\}$;
10: **while** $\mathcal{NE}_{d|Q}^{\preceq}(U') > \mathcal{NE}_{d|A}^{\preceq}(U')$ **do**
11: **for** $h = 1$ to $|A - Q|$ **do**
12: select $Q \leftarrow Q \cup \{a'_h\}$ and calculate $\mathcal{NE}_{d|Q}^{\preceq}(U')$;
13: **end for**
14: **end while**
15: **for** each $a \in Q$ **do**
16: compute FDNCE $\mathcal{NE}_{d|(Q-\{a\})}^{\preceq}(U')$ via using Eq. (26);
17: **if** $\mathcal{NE}_{d|(Q-\{a\})}^{\preceq}(U') \leq \mathcal{NE}_{d|Q}^{\preceq}(U')$ **then**
18: $Q \leftarrow Q - \{a\}$;
19: **end if**
20: **end for**
21: $Red_{U'} \leftarrow Q$;
22: **return** $Red_{U'}$;

1) FDNCE-IFSD algorithm (see Algorithm 3): In Algorithm 3, Step 1 is to delete the object set. Step 2 is to update the original diagonal matrices by Proposition 2. Step 3 is to compute the new FDNCE via using Eq. (26). Steps 4-8 is to determine whether the new FDNCE under the original reduct is not higher than that of under the entire attribute set, if so, then keep the original reduct unchanged. Steps 9-14 is to construct a descending sequence for the remaining attributes, and incrementally update the selected feature subset until Step 10 does not hold. Steps 15-20 is to remove redundant attributes from the selected attribute subset. Steps 21-22 is to output the final reduct.

2) The time complexity of FDNCE-IFSD algorithm: The time complexity of the main steps in this algorithm are listed in Table VI.

3) The comparison of time complexity: The time complexity of algorithms FDNCE-HFS and FDNCE-IFSD are shown in Table VII for intuitive comparison. From Table VII, obviously, the time complexity of FDNCE-IFSD algorithm is much lower than that of FDNCE-HFS algorithm. The main reason is that FDNCE-IFSD algorithm uses the previous knowledge when calculating the new reduct, while FDNCE-HFS algorithm calculates a new reduct from scratch, which does not use the previous knowledge. So FDNCE-HFS algorithm is very time consuming for calculating a new reduct.

TABLE VI
THE TIME COMPLEXITY OF FDNCE-IFSD ALGORITHM

Steps	Time complexity	Steps	Time complexity
2-3	$O(\lvert U_{de}\rvert\lvert U\rvert)$	15-20	$O(\lvert Q\rvert^2\lvert U'\rvert^2)$
9-14	$O((\lvert A\rvert - \lvert Q\rvert)\lvert U'\rvert^2)$		

TABLE VII
THE COMPARISON OF THE TIME COMPLEXITY OF ALGORITHMS FDNCE-HFS AND FDNCE-IFSD

Algorithms	Time complexity
FDNCE-HFS	$O(\lvert A\rvert\lvert U'\rvert^2 + \lvert A\rvert^2\lvert U'\rvert^2 + \lvert A\rvert^2\lvert U'\rvert^2 + \lvert Q\rvert^2\lvert U'\rvert^2)$
FDNCE-IFSD	$O(\lvert U_{de}\rvert\lvert U\rvert + (\lvert A\rvert - \lvert Q\rvert)\lvert U'\rvert^2 + \lvert Q\rvert^2\lvert U'\rvert^2)$

VI. EXPERIMENTS AND ANALYSIS

In this section, we perform a series of experiments to test the robustness of the proposed metric and evaluate the performance of the proposed feature selection algorithms. The configuration of computer used for experiments is as follows. CPU is Intel(R) Core(TM) i7-8700. Clock Speed is 3.20 GHz. Memory is 16.0 GB. Operation System is 64-bit Windows 10. The algorithms are coded by Java. We downloaded ten datasets from the UCI machine learning repository, and a summary of them is given in Table VIII.

TABLE VIII
THE SUMMARY OF DATASETS

No.	Datasets	Abbreviation	Samples	Features	Classes
1	Wisconsin Prognostic Breast Cancer	WPBC	198	32	2
2	Dermatology	Derm	358	34	6
3	Libras Movement	Libras	360	90	15
4	Australian Credit	Aust	690	14	2
5	German Credit	Germ	1000	20	2
6	Mice Protein Expression	Mice	1077	68	8
7	Car Evaluation	Car	1728	6	4
8	Cardiotocography	Card	2126	21	3
9	Waveform	Wave	5000	21	3
10	Nursery	Nurs	8029	8	5

Before conducting the experiments, we preprocess these datasets. For categorical features, we use integers instead of symbols, and define order relation of the integers in accordance with semantics of the features. These datasets are normalized via using

$$\hat{v}_{ik} = \frac{v_{ik} - min(V_{a_k})}{max(V_{a_k}) - min(V_{a_k})}. \tag{29}$$

These preprocessed datasets are saved in the GitHub homepage [1].

To evaluate the effectiveness of feature selection algorithms, two classifiers K-nearest neighbor (KNN, K=3) and support vector machine (SVM) are applied to the datasets after reduction to verify the effectiveness of feature selection methods.

[1] https://github.com/binbinsang/Incremental-FS-FDNRS-dataset-R1.git

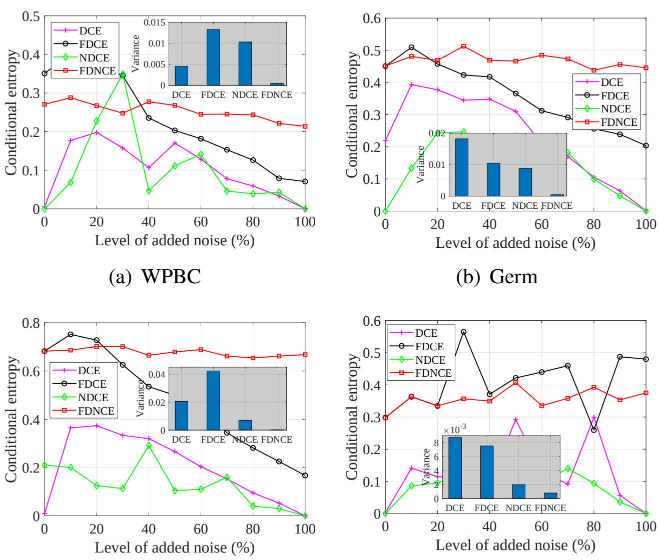
10-fold cross-validation is adopted in classification. The experimental process is repeated 10 rounds on each dataset, and the mean and standard deviation of classification accuracy are recorded and compared. For dynamic data, the reduct obtained by running the feature selection algorithm may be different in different runs. Therefore, the average of reduct sizes in ten runs is adopted as the reduct size.

A. The robustness evaluations of metric FDNCE

In this subsection, we randomly select four datasets in Table VIII to test the robustness of metrics DCE, FDCE, NDCE, and FDNCE. For each dataset, we choose different proportions of data to add random noise. These datasets with noise are obtained via using

$$\hat{v}_{ij} = \begin{cases} \hat{v}_{ij} + r_{ij}, & 0 \leq \hat{v}_{ij} + r_{ij} \leq 1; \\ \hat{v}_{ij}, & otherwise, \end{cases} \quad (30)$$

where $r_{ij} \in [0, 1]$. Then, these four metrics are calculated for different levels noise datasets. The experimental results are presented in Fig. 4, where the histogram in each subgraph shows the variance of the conditional entropy under different noise levels.



(a) WPBC (b) Germ

(c) Mice (d) Card

Fig. 4. The comparison of robustness of metrics at different noise levels

Fig. 4 indicates that the fluctuation of FDNCE curve is relatively small as the noise level increases. Moreover, in each sub-figure, we also show the variance of the calculation result of each metric. From these histograms, we can intuitively observe that the variance of FDNCE is the minimum one. Therefore, we can conclude that the robustness of metric FDNCE is the best one compared with other three metrics.

B. The effectiveness evaluations of FDNCE-HFS algorithm

This subsection compares the classification performance of the reducts obtained via HFS based on DCE, NDCE, FDCE, and FDNCE, respectively. Table IX shows the results of the experiment, where "raw" is the classification accuracy of the raw feature set. Note that in Table IX, the number in bracket after each classification accuracy result indicates the size of the generated reduct. In the following subsections, the structure of Tables X and XI is similar to Table IX.

From Table IX, it show that the classification accuracy of the reducts obtained via FDNCE-HFS algorithm in most datasets is not only higher than that of the raw feature set, but also higher than that of HFS algorithm using the other three metrics. The average value of classification accuracy of FDNCE-HFS algorithm is the highest one. Hence, the reduct generated by using FDNCE-HFS algorithm is better. It is conclude that FDNCE-HFS algorithm can precisely remove redundant attributes in ordered data and improve classification performance.

TABLE X
THE CLASSIFICATION ACCURACY OF GENERATED REDUCT VIA USING DIFFERENT ALGORITHMS (%)

Datasets	KNN		SVM	
	FDNCE-HFS	FDNCE-IFSA	FDNCE-HFS	FDNCE-IFSA
WPBC	52.6±1.6 (14.0)	45.9±2.1 (7.3)	53.5±1.8 (14.0)	59.8±1.7 (7.3)
Derm	94.1±0.6 (10.0)	94.3±0.4 (10.0)	97.6±0.3 (10.0)	97.4±0.2 (10.0)
Libras	88.5±0.8 (36.0)	89.1±0.8 (35.6)	70.6±1.1 (36.0)	71.9±1.1 (35.6)
Aust	77.0±0.7 (8.0)	76.5±1.0 (7.4)	85.3±0.2 (8.0)	85.5±0.1 (7.4)
Germ	62.0±0.5 (5.0)	63.8±0.7 (6.2)	70.1±0.1 (5.0)	70.0±0.1 (6.2)
Mice	91.2±0.5 (6.0)	91.9±0.6 (6.0)	92.9±0.2 (6.0)	93.0±0.4 (6.0)
Car	90.8±0.2 (5.0)	90.7±0.3 (5.0)	85.6±0.3 (5.0)	85.5±0.1 (5.0)
Card	91.0±0.2 (8.0)	84.7±0.3 (3.0)	90.1±0.1 (8.0)	81.9±0.2 (3.0)
Wave	75.7±0.2 (14.0)	75.6±0.2 (14.2)	84.3±0.1 (14.0)	84.3±0.1 (14.2)
Nurs	89.0±0.1 (5.0)	89.0±0.1 (5.0)	90.2±0.1 (5.0)	90.2±0.1 (5.0)
Average	81.2±0.6 (11.1)	80.1±0.6 (10.0)	82.0±0.4 (11.1)	81.9±0.4 (10.0)

[1] The size of the reduct is the average of the reducts generated by running the algorithm ten times.

C. The performance evaluations of FDNCE-IFSA algorithm

In this subsection, we evaluate the performance of algorithm FDNCE-IFSA in terms of effectiveness and efficiency. In terms of effectiveness, we compare algorithms FDNCE-IFSA and FDNCE-HFS from two aspects: reduct size and its classification performance. In terms of efficiency, we compare algorithms FDNCE-IFSA and FDNCE-HFS from two aspects: computational time and speed-up ratio.

1) Effectiveness evaluations: The dynamic datasets are simulated by the following way. For each preprocessed dataset, 50% of the objects are randomly sampled as an initial object set U, and the all remaining objects are treated as an added object set U_{ad}. Algorithms FDNCE-IFSA and FDNCE-HFS are conducted to obtain a new reduct when U_{ad} is added to U. Then, the classification accuracy of the reducts obtained by these two algorithms are verified and compared. The experimental results are presented in Table X.

From Table X, we can see that the classification performance of the reducts obtained by algorithms FDNCE-IFSA and FDNCE-HFS is almost equal in most datasets. Moreover, the size of the reducts generated by these two algorithms are equal or very close in most datasets. This finding proves that the reducts obtained by algorithms FDNCE-IFSA and FDNCE-HFS have almost the same classification performance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

TABLE IX
THE CLASSIFICATION ACCURACY OF REDUCTS OBTAINED VIA ALGORITHM HFS WITH DIFFERENT METRICS (%)

Datasets	KNN					SVM				
	Raw	DCE	NDCE	FDCE	FDNCE	Raw	DCE	NDCE	FDCE	FDNCE
WPBC	50.9±1.6	48.4±2.3 (10.0)	47.3±1.3 (9.0)	51.9±2.4 (16.0)	**52.6±1.6 (14.0)**	53.6±1.8	**57.6±2.3 (10.0)**	57.3±2.1 (9.0)	52.8±4.4 (16.0)	53.5±1.8 (14.0)
Derm	89.9±1.0	86.5±1.0 (12.0)	53.4±0.6 (2.0)	75.5±0.4 (6.0)	**94.1±0.6 (10.0)**	95.7±0.7	90.1±0.2 (12.0)	55.4±0.1 (2.0)	77.0±0.4 (6.0)	**97.6±0.3 (10.0)**
Libras	87.1±0.6	78.2±0.6 (14.0)	76.5±0.6 (15.0)	88.1±0.8 (37.0)	**88.5±0.8 (36.0)**	70.2±1.3	59.8±1.3 (14.0)	56.9±1.2 (15.0)	62.5±1.1 (37.0)	**70.6±1.1 (36.0)**
Aust	65.2±0.6	69.5±0.6 (11.0)	69.2±0.4 (4.0)	77.0±0.7 (8.0)	**77.0±0.7 (8.0)**	73.6±1.5	69.8±5.8 (11.0)	75.5±0.1 (4.0)	85.3±0.2 (8.0)	**85.3±0.2 (8.0)**
Germ	60.9±0.9	60.4±0.9 (18.0)	57.7±0.5 (4.0)	61.4±0.5 (5.0)	**62.0±0.5 (5.0)**	58.1±3.8	56.3±3.6 (18.0)	69.6±0.2 (4.0)	69.8±0.3 (5.0)	**70.1±0.1 (5.0)**
Mice	**99.5±0.1**	89.5±0.2 (20.0)	88.5±0.2 (20.0)	89.5±0.4 (4.0)	91.2±0.5 (6.0)	**93.8±0.3**	78.4±0.5 (20.0)	85.4±0.1 (20.0)	90.4±0.2 (4.0)	92.9±0.2 (6.0)
Car	87.3±0.2	87.3±0.2 (6.0)	81.3±0.2 (5.0)	90.8±0.2 (5.0)	**90.8±0.2 (5.0)**	85.1±0.1	85.1±0.1 (6.0)	81.1±0.1 (5.0)	85.6±0.3 (5.0)	**85.6±0.3 (5.0)**
Card	90.7±0.3	89.4±0.2 (12.0)	71.8±0.1 (2.0)	91.0±0.2 (8.0)	**91.0±0.2 (8.0)**	88.3±0.2	85.1±0.1 (12.0)	78.2±0.1 (2.0)	89.1±0.1 (8.0)	**90.1±0.1 (8.0)**
Wave	**77.3±0.3**	76.0±0.2 (16.0)	76.6±0.2 (16.0)	75.0±0.2 (13.0)	75.7±0.2 (14.0)	**86.9±0.1**	85.1±0.1 (16.0)	85.6±0.1 (16.0)	84.0±0.1 (13.0)	84.3±0.1 (14.0)
Nurs	**92.4±0.2**	44.3±0.2 (7.0)	44.5±0.1 (7.0)	65.2±0.1 (4.0)	89.0±0.1 (5.0)	88.8±0.1	53.7±0.1 (7.0)	53.5±0.1 (7.0)	75.2±0.1 (4.0)	**90.2±0.1 (5.0)**
Average	80.1±0.6	73.0±0.6 (12.6)	66.7±0.4 (8.4)	76.5±0.6 (10.6)	**81.2±0.6 (11.1)**	79.4±1.0	72.1±1.4 (12.6)	69.8±0.4 (8.4)	77.2±0.7 (10.6)	**82.0±0.4 (11.1)**

Hence, we can conclude from Table X that FDNCE-IFSA algorithm is effective.

2) Efficiency evaluations: In the previous experiment, we have divided each preprocessed data into the initial object set U and the added object set U_{ad}. The dynamic change of datasets is simulated in the following way. Different ratios of objects sampled randomly from U_{ad} are added to U to obtain dynamic datasets for testing ($i.e.$, 10%, 20%, 30%, 40%, and 50% of the objects from U_{ad} are randomly sampled and added to U). The time consumption of algorithms FDNCE-IFSA and FDNCE-HFS are compared by using dynamic testing sets. The experimental results are presented in Fig. 5.

From Fig. 5, each sub-figure shows that the computational time of FDNCE-IFSA algorithm is remarkably less than that of FDNCE-HFS algorithm. Furthermore, as the size of the added object set increases, the growth trend of the time consumed via FDNCE-IFSA algorithm is slower than that via FDNCE-HFS algorithm. For datasets Derm, Libras, and Mice with larger feature numbers, the time-consuming of the incremental algorithm is also significantly lower than that of the non-incremental algorithm. Moreover, for datasets Wave and Nurs with larger sample numbers, the computational efficiency of the incremental algorithm is also observably higher than that of the non-incremental algorithm. This finding proves that FDNCE-IFSA algorithm can efficiently obtain a reduct when adding objects. In particular, compared with non-incremental algorithms, its computational efficiency is not affected by the feature set and sample set size of the dataset.

Subsequently, we again demonstrate the efficiency of FDNCE-IFSA algorithm again by speed-up ratio, which is calculated as $S = T_{FDNCE-HFS}/T_{FDNCE-IFSA}$, T_* is the computational time of $*$ algorithm. Based on the results shown in Fig. 5, the speed-up ratio of each dataset is calculated. The experimental results are shown in Fig. 6.

As shown in Fig. 6, algorithm FDNCE-IFSAis at least nearly two times or more faster than FDNCE-HFS algorithm on all the datasets except dataset Car. It is worth pointing out that for dataset Mice with larger features, FDNCE-IFSA algorithm is at least six times faster than FDNCE-HFS algorithm, and for datasets Wave with larger sample set, FDNCE-IFSA algorithm is approximately four times faster than FDNCE-HFS algorithm. The experimental results again prove that the efficiency of FDNCE-IFSA algorithm.

3) Summary: From the evaluations of effectiveness and efficiency of FDNCE-IFSA algorithm, a conclusion can be drawn that the computational time required to obtain a feasible reduct via FDNCE-IFSA algorithm is considerably shorter than that required via FDNCE-HFS algorithm. Therefore, when adding multiple objects to an ODS, the proposed incremental algorithm FDNCE-IFSA can efficiently generate a feasible reduct without reducing the classification performance.

D. The performance evaluations of algorithm FDNCE-IFSD

This subsection evaluates the performance of FDNCE-IFSD algorithm in terms of effectiveness and efficiency. Algorithms FDNCE-IFSD and FDNCE-HFS are compared in the same scheme as the previous subsection.

1) Effectiveness evaluations: The dynamic datasets are simulated in the following way. Naturally, each preprocessed dataset is taken as an initial object set U, and then 50% of the objects are randomly sampled as a deleted object set U_{de}. Algorithms FDNCE-IFSD and FDNCE-HFS are used to calculate a new reduct when objects are deleted. Then, the classification accuracy of the reducts obtained by these two algorithms is compared. The experimental results are presented in Table XI.

TABLE XI
THE CLASSIFICATION ACCURACY OF GENERATED REDUCT VIA DIFFERENT ALGORITHMS (%)

Datasets	KNN		SVM	
	FDNCE-HFS	FDNCE-IFSD	FDNCE-HFS	FDNCE-IFSD
WPBC	51.0±2.4 (15.8)	50.1±1.2 (13.3)	55.3±2.8 (15.8)	56.9±3.1 (13.3)
Derm	92.2±0.8 (12.4)	92.1±0.5 (9.6)	95.9±0.3 (12.4)	95.1±0.7 (9.6)
Libras	89.1±1.5 (46.8)	89.0±1.4 (36.9)	68.4±2.1 (46.8)	70.1±2.6 (36.9)
Aust	78.9±1.1 (6.0)	78.4±0.9 (7.2)	85.5±0.1 (6.0)	85.2±0.3 (7.2)
Germ	66.0±0.7 (12.0)	65.5±1.0 (5.0)	56.5±5.2 (12.0)	58.6±3.6 (5.0)
Mice	92.0±1.5 (6.2)	93.5±2.3 (5.7)	91.9±1.5 (6.2)	92.1±1.2 (5.7)
Car	91.9±0.1 (4.0)	94.5±0.3 (5.0)	89.1±0.3 (4.0)	88.6±0.3 (5.0)
Card	88.4±2.9 (4.0)	88.6±2.2 (4.4)	85.2±0.1 (4.0)	85.1±0.3 (4.4)
Wave	74.7±0.3 (13.9)	74.9±0.3 (14.2)	83.0±0.3 (13.9)	83.2±0.1 (14.2)
Nurs	84.0±0.1 (5.0)	84.0±0.1 (5.0)	90.4±0.1 (5.0)	90.3±0.1 (5.0)
Average	80.8±1.1 (12.6)	81.1±1.0 (10.6)	80.1±1.3 (12.6)	80.5±1.2 (10.6)

[1] The size of the reduct is the average of the reducts generated by running the algorithm ten times.

From Table XI, we find that the size of the reducts generated by these two algorithms are equal or very close in most datasets. It is worth noting that the classification performance

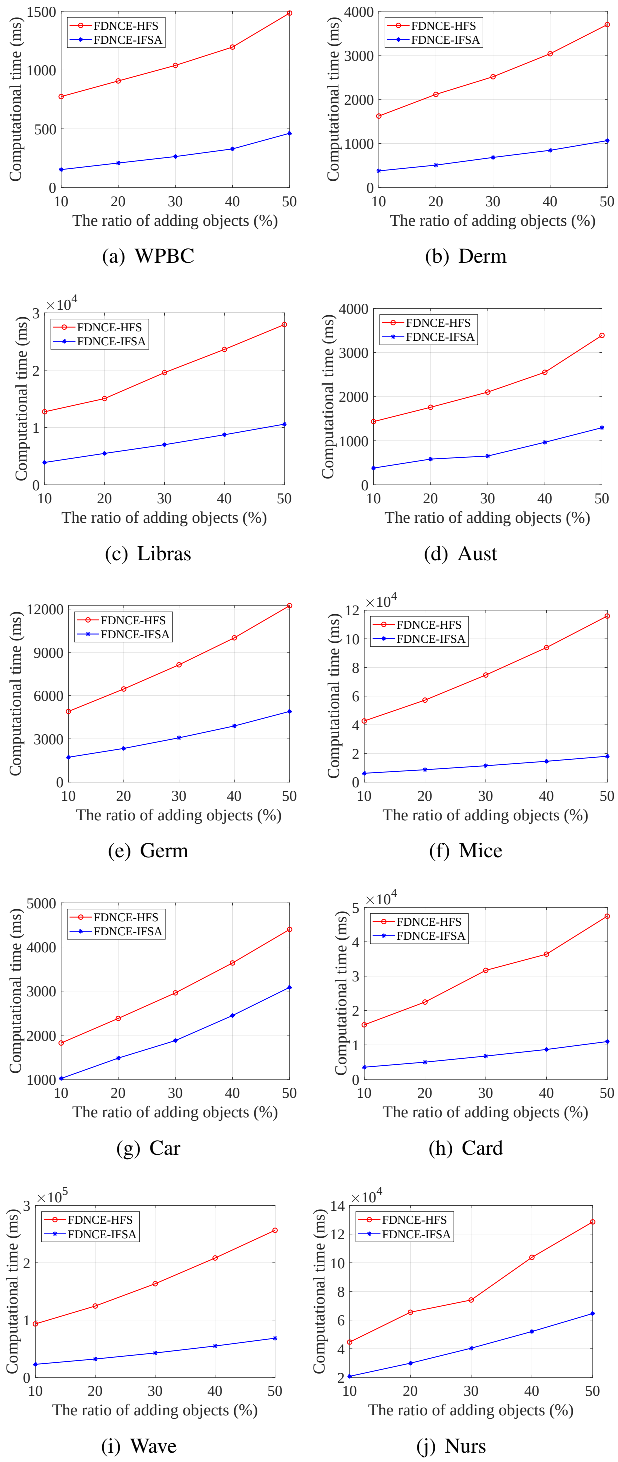


Fig. 5. The computational time of different algorithms versus different ratios of adding objects

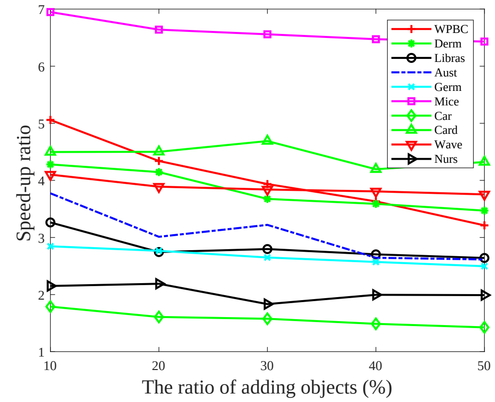


Fig. 6. The speed-up ratio of algorithm FDNCE-IFSA

of the reducts obtained by algorithms FDNCE-IFSD and FDNCE-HFS is nearly equal in most datasets. This finding proves that the reducts obtained by algorithms FDNCE-IFSD and FDNCE-HFS have almost the same classification performance. Hence, the experimental results indicate that algorithm FDNCE-IFSD is effective.

2) Efficiency evaluations: The dynamic change of datasets is simulated in the following way. For each preprocessed dataset, different ratios of objects are randomly sampled from the initial object set U as deleting objects (*i.e.*, 10%, 20%, 30%, 40%, and 50% of U are respectively deleted to construct testing sets). Then, the running time of algorithms FDNCE-IFSD and FDNCE-HFS on testing sets are recorded. The change trend lines of these two algorithms are shown in Fig. 7.

Fig. 7 clearly shows that as the size of deleted object set increases, the running time of algorithms FDNCE-IFSD and FDNCE-HFS decreases. Notably, the running time of FDNCE-IFSD algorithm is remarkably less than that of FDNCE-HFS algorithm. This proves that FDNCE-IFSD algorithm is more efficient than FDNCE-HFS algorithm. It is worth noting that for datasets Derm, Libras, and Mice with large feature scales, the time cost of algorithm FDNCE-IFSD is much lower than that of algorithm FDNCE-HFS. Furthermore, for datasets Wave and Nurs with a large sample set, the time-consuming of algorithm FDNCE-IFSD is also significantly lower than that of algorithm FDNCE-HFS. In addition, we can conclude from the above two points that the computational efficiency of the incremental algorithm FDNCE-IFSD does not change linearly with the size of the feature set or sample set.

Afterwards, the efficiency of FDNCE-IFSD algorithm is verified again by calculating the speed-up ratio of the running algorithms. Similarly, the speed-up ratio of each dataset is calculated according to the results in Fig. 7. The results of the experiment are shown in Fig. 8.

Fig. 8 indicates that FDNCE-IFSD algorithm is at least nearly two times or more faster than FDNCE-HFS algorithm for all datasets. Especially for datasets Mice with larger feature numbers, algorithm FDNCE-IFSD is at least ten times faster than algorithm FDNCE-HFS, and for datasets Wave with larger sample set, algorithm FDNCE-IFSD is at least four times faster than FDNCE-HFS algorithm. The experimental results again testify that FDNCE-IFSD algorithm has higher efficiency than FDNCE-HFS algorithm.

3) Summary: After experimental analysis, it can be concluded that FDNCE-IFSD algorithm not only decreases the computational time, but also does not lessen the classification performance. Accordingly, compared with FDNCE-HFS

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

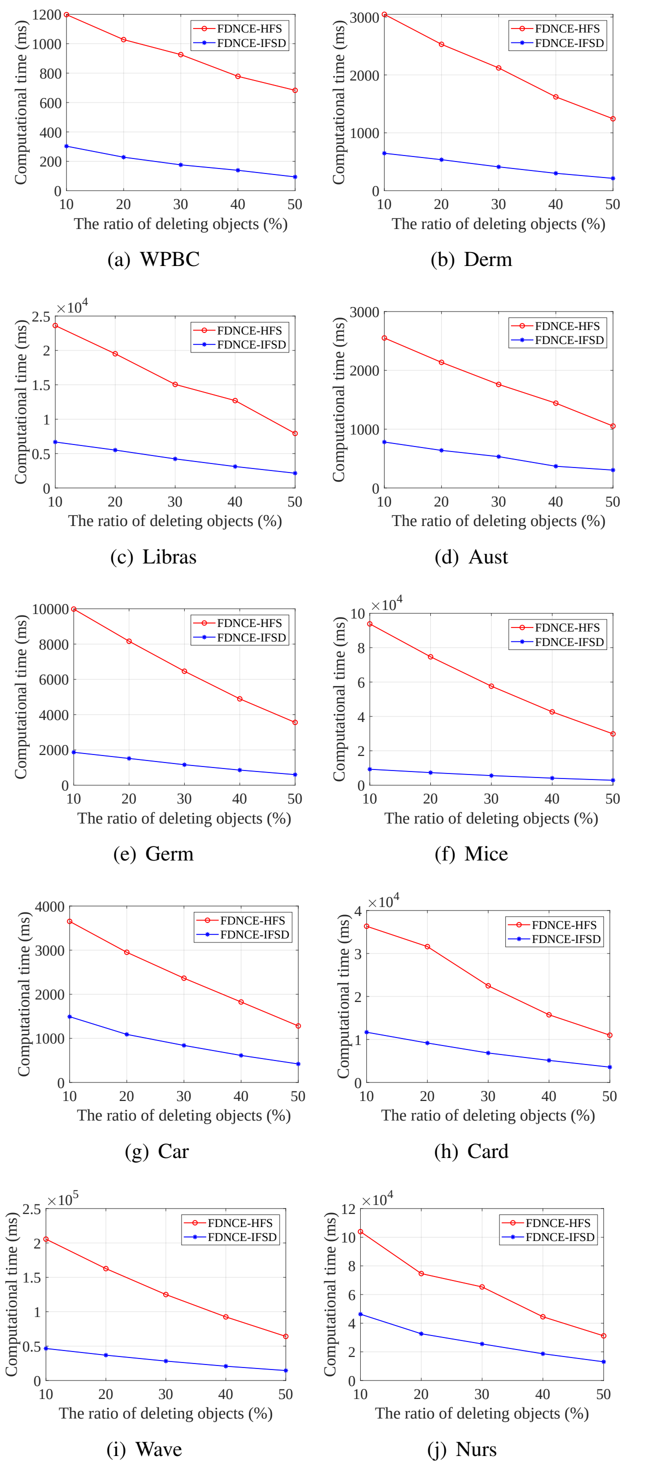


Fig. 7. The computational time of different algorithms versus different ratios deleting objects

algorithm, FDNCE-IFSD algorithm can quickly generate a satisfying reduct when deleting multiple objects from an ODS.

VII. Conclusion and Future Work

Feature selection is an effective information preprocessing technology, which can effectively remove redundant attributes and improve classification performance. However, with the development of the information age, different types of data have different requirements for feature selection methods. This study investigate incremental feature selection approaches for dynamic ordered data with time-evolving objects under FDNRS model framework. Experiments are performed on ten public datasets. The findings from the experimental results are: (1) The metric FDNCE is more robust for ordered data with noise. (2) The classification ability of the reducts obtained via FDNCE-HFS algorithm is not only higher than that of the raw feature set, but also higher than that of HFS algorithm using other metrics. (3) The proposed incremental feature selection algorithms can efficiently calculate an effective reduct from dynamic ordered data with time-evolving objects.

In this study, the developed incremental feature selection approaches are suitable for dynamic ordered data with the variation of objects. Nevertheless, dynamic ordered data with the variation of multi-sided is closer to reality, which inspire our further research. In future work, based on the current research results, we will investigate incremental feature selection approaches for dynamic ordered data with the variation of multi-sided.

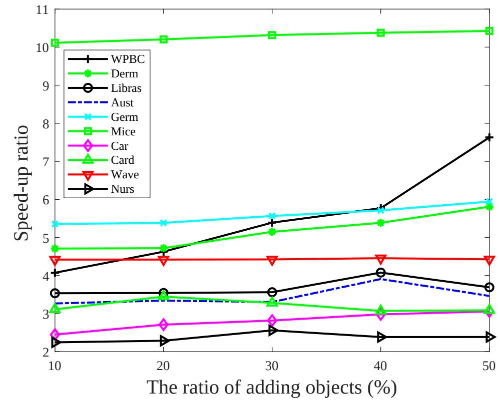


Fig. 8. The speed-up ratio of algorithm FDNCE-IFSD

References

[1] W. P. Ding, C. T. Lin, and W. Pedrycz, "Multiple relevant feature ensemble selection based on multilayer co-evolutionary consensus mapreduce," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 425 – 439, 2020.

[2] W. P. Ding, C. T. Lin, and Z. H. Cao, "Shared nearest-neighbor quantum game-based attribute reduction with hierarchical coevolutionary spark and its application in consistent segmentation of neonatal cerebral cortical surfaces," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 7, pp. 2013 – 2027, 2019.

[3] Q. H. Hu, L. J. Zhang, Y. C. Zhou, and W. Pedrycz, "Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 226 – 238, 2018.

[4] L. Zhao, Q. H. Hu, and W. W. Wang, "Heterogeneous feature selection with multi-modal deep neural networks and sparse group lasso," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1936 – 1948, 2015.

[5] Y. J. Lin, Q. H. Hu, J. H. Liu, J. J. Li, and X. D. Wu, "Streaming feature selection for multilabel learning based on fuzzy mutual information," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1491 – 1507, 2017.

[6] J. Y. Liang, F. Wang, C. Y. Dang, and Y. H. Qian, "A group incremental approach to feature selection applying rough set technique," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 294 – 308, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3064686, IEEE Transactions on Fuzzy Systems

[7] D. G. Chen, Y. Y. Yang, and Z. Dong, "An incremental algorithm for attribute reduction with variable precision rough sets," *Applied Soft Computing*, vol. 45, pp. 129 – 149, 2016.

[8] Y. Y. Yang, D. G. Chen, H. Wang, E. C. Tsang, and D. L. Zhang, "Fuzzy rough set based incremental attribute reduction from dynamic data with sample arriving," *Fuzzy Sets and Systems*, vol. 312, pp. 66 – 86, 2017.

[9] K. Gong, Y. Wang, M. Z. Xu, and Z. Xiao, "Bssreduce an o(vertical bar u vertical bar) incremental feature selection approach for large-scale and high-dimensional data," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3356 – 3367, 2018.

[10] W. H. Shu, W. B. Qian, and Y. H. Xie, "Incremental approaches for feature selection from dynamic data with the variation of multiple objects," *Knowledge-Based Systems*, vol. 163, pp. 320 – 331, 2019.

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

[12] J. H. Dai, H. Hu, W. Z. Wu, Y. H. Qian, and D. B. Huang, "Maximal discernibility pairs based approach to attribute reduction in fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 2174–2187, 2018.

[13] J. H. Dai, Q. H. Hu, H. Hu, and D. B. Huang, "Neighbor inconsistent pair selection for attribute reduction by rough set approach," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 937 – 950, 2018.

[14] C. Z. Wang, Y. Wang, M. W. Shao, Y. H. Qian, and D. G. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 818 – 830, 2020.

[15] L. Sun, X. Y. Zhang, Y. H. Qian, J. C. Xu, and S. G. Zhang, "Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification," *Information Sciences*, vol. 502, pp. 18 – 41, 2019.

[16] S. Greco, B. Matarazzo, and R. Slowinski, "Rough approximation of a preference relation by dominance relations," *European Journal of Operational Research*, vol. 117, no. 1, pp. 63–83, 1999.

[17] S. Greco, B. Matarazzo, and R. Slowinski, "Rough sets theory for multicriteria decision analysis," *European Journal of Operational Research*, vol. 129, no. 1, pp. 1 – 47, 2001.

[18] H. M. Chen, T. R. Li, Y. Cai, C. Luo, and H. Fujita, "Parallel attribute reduction in dominance-based neighborhood rough set," *Information Sciences*, vol. 373, pp. 351 – 368, 2016.

[19] Q. H. Hu, D. R. Yu, and M. Z. Guo, "Fuzzy preference based rough sets," *Information Sciences*, vol. 180, no. 10, pp. 2003 – 2022, 2010.

[20] C. Shannon and W. Weaver, "The mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3/4, pp. 373 – 423, 1948.

[21] Q. H. Hu, M. Z. Guo, D. R. Yu, and J. F. Liu, "Information entropy for ordinal classification," *Science China Information Sciences*, vol. 53, no. 6, pp. 1188–1200, 2010.

[22] Q. H. Hu, W. W. Pan, L. Zhang, D. Zhang, Y. P. Song, M. Z. Guo, and D. R. Yu, "Feature selection for monotonic classification," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 69 – 81, 2012.

[23] Q. H. Hu, X. J. Che, L. Zhang, D. Zhang, M. Z. Guo, and D. R. Yu, "Rank entropy based decision trees for monotonic classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2052 – 2064, 2012.

[24] R. Susmaga, "Reducts and constructs in classic and dominance-based rough sets approach," *Information Sciences*, vol. 271, pp. 45 – 64, 2014.

[25] H. Y. Zhang and S. Y. Yang, "Feature selection and approximate reasoning of large-scale set-valued decision tables based on $\alpha-$dominance-based quantitative rough sets," *Information Sciences*, vol. 378, pp. 328 – 347, 2017.

[26] W. B. Qian and W. H. Shu, "Attribute reduction in incomplete ordered information systems with fuzzy decision," *Applied Soft Computing*, vol. 73, pp. 242 – 253, 2018.

[27] W. S. Du and B. Q. Hu, "A fast heuristic attribute reduction approach to ordered decision systems," *European Journal of Operational Research*, vol. 264, no. 2, pp. 440 – 452, 2018.

[28] J. H. Yu, M. H. Chen, and W. H. Xu, "Dynamic computing rough approximations approach to time-evolving information granule interval-valued ordered information system," *Applied Soft Computing*, vol. 60, pp. 18 – 29, 2017.

[29] X. Yang, T. R. Li, D. Liu, and H. Fujita, "A multilevel neighborhood sequential decision approach of three-way granular computing," *Information Sciences*, vol. 538, pp. 119 – 141, 2020.

[30] Y. J. Zhang, D. Q. Miao, W. Pedrycz, T. N. Zhao, J. F. Xu, and Y. Yu, "Granular structure-based incremental updating for multi-label classification," *Knowledge-Based Systems*, vol. 189, pp. 1 – 15, 2020.

[31] S. Bouzayane and I. Saad, "A multicriteria approach based on rough set theory for the incremental periodic prediction," *European Journal of Operational Research*, vol. 286, no. 1, pp. 282 – 298, 2020.

[32] X. Zhang, C. L. Mei, D. G. Chen, Y. Y. Yang, and J. H. Li, "Active incremental feature selection using a fuzzy rough set-based information entropy," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 901 – 915, 2020.

[33] N. L. Giang, L. H. Son, T. T. Ngan, T. M. Tuan, H. T. Phuong, M. Abdel-Basset, A. R. L. de Macedo, and V. H. C. de Albuquerque, "Novel incremental algorithms for attribute reduction from dynamic decision tables using hybrid filtercwrapper with fuzzy partition distance," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 858 – 873, 2020.

[34] Y. Y. Yang, D. G. Chen, and W. Hui, "Active sample selection based incremental algorithm for attribute reduction with rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 825 – 838, 2017.

[35] Y. Y. Yang, D. G. Chen, H. Wang, and X. Z. Wang, "Incremental perspective for feature selection based on fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1257 – 1273, 2018.

[36] Y. Y. Yang, S. J. Song, D. G. Chen, and X. Zhang, "Discernible neighborhood counting based incremental feature selection for heterogeneous data," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 5, pp. 1115 – 1127, 2020.

[37] W. H. Shu, W. B. Qian, and Y. H. Xie, "Incremental feature selection for dynamic hybrid data using neighborhood rough set," *Knowledge-Based Systems*, vol. 194, pp. 1 – 15, 2020.

[38] Y. Liu, L. D. Zheng, Y. L. Xiu, H. Yin, S. Y. Zhao, X. Z. Wang, H. Chen, and C. P. Li, "Incremental feature selection for dynamic hybrid data using neighborhood rough set," *International Journal of approximate reasoning*, vol. 118, pp. 1 – 26, 2020.

[39] A. K. Das, S. Sengupta, and S. Bhattacharyya, "A group incremental feature selection for classification using rough set theory based genetic algorithm," *Applied Soft Computing*, vol. 65, pp. 400 – 411, 2018.

[40] B. B. Sang, H. M. Chen, T. R. Li, W. H. Xu, and H. Yu, "Incremental approaches for heterogeneous feature selection in dynamic ordered data," *Information Sciences*, vol. 541, pp. 475 – 501, 2020.

[41] P. Ni, S. Y. Zhao, X. Z. Wang, H. Chen, C. P. Li, and E. C. Tsang, "Incremental feature selection based on fuzzy rough sets," *Information Sciences*, vol. 536, pp. 185 – 204, 2020.

[42] D. G. Chen, L. J. Dong, and J. S. Mi, "Incremental mechanism of attribute reduction based on discernible relations for dynamically increasing attribute," *Soft Computing*, vol. 24, no. 1, pp. 321 – 332, 2020.

[43] F. Wang, J. Y. Liang, and Y. H. Qian, "Attribute reduction: A dimension incremental strategy," *Knowledge-Based Systems*, vol. 39, pp. 95 – 108, 2013.

[44] G. M. Lang, D. Q. Miao, M. J. Cai, and Z. F. Zhang, "Incremental approaches for updating reducts in dynamic covering information systems," *Knowledge-Based Systems*, vol. 134, pp. 85 – 104, 2017.

[45] A. P. Zeng, T. R. Li, D. Liu, J. B. Zhang, and H. M. Chen, "A fuzzy rough set approach for incremental feature selection on hybrid information systems," *Fuzzy Sets and Systems*, vol. 258, pp. 39 – 60, 2015.

[46] W. Wei, X. Y. Wu, J. Y. Liang, J. B. Cui, and Y. J. Sun, "Discernibility matrix based incremental attribute reduction for dynamic data," *Knowledge-Based Systems*, vol. 140, pp. 142 – 157, 2018.

[47] W. Wei, P. Song, J. Y. Liang, and X. Y. Wu, "Accelerating incremental attribute reduction algorithm by compacting a decision table," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 9, pp. 2355 – 2373, 2019.

[48] M. J. Cai, G. M. Lang, H. Fujita, Z. Y. Li, and T. Yang, "Incremental approaches to updating reducts under dynamic covering granularity," *Knowledge-Based Systems*, vol. 172, pp. 130 – 140, 2019.

[49] L. J. Dong and D. G. Chen, "Incremental attribute reduction with rough set for dynamic datasets with simultaneously increasing samples and attributes," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 6, pp. 1339 – 1355, 2020.

[50] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 111 – 127, 1997.