Computer Systems
Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.042068

ARTICLE

A Novel Incremental Attribute Reduction Algorithm Based on Intuitionistic Fuzzy Partition Distance

Pham Viet Anh[1,3], Nguyen Ngoc Thuy[4], Nguyen Long Giang[2], Pham Dinh Khanh[5] and Nguyen The Thuy[1,6,*]

[1]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, 100000, Vietnam

[2]Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, 100000, Vietnam

[3]HaUI Institute of Technology, Hanoi University of Industry, Hanoi, 100000, Vietnam

[4]Faculty of Information Technology, University of Sciences, Hue University, Hue, 530000, Vietnam

[5]AI Research Department, Neurond Technology JSC, Hanoi, 100000, Vietnam

[6]Information and Communication Technology Center, Department of Information and Communications, Bacninh, 790000, Vietnam

*Corresponding Author: Nguyen The Thuy. Email: thethuy@gmail.com

Received: 17 May 2023 Accepted: 05 July 2023 Published: 09 November 2023

ABSTRACT

Attribute reduction, also known as feature selection, for decision information systems is one of the most pivotal issues in machine learning and data mining. Approaches based on the rough set theory and some extensions were proved to be efficient for dealing with the problem of attribute reduction. Unfortunately, the intuitionistic fuzzy sets based methods have not received much interest, while these methods are well-known as a very powerful approach to noisy decision tables, i.e., data tables with the low initial classification accuracy. Therefore, this paper provides a novel incremental attribute reduction method to deal more effectively with noisy decision tables, especially for high-dimensional ones. In particular, we define a new reduct and then design an original attribute reduction method based on the distance measure between two intuitionistic fuzzy partitions. It should be noted that the intuitionistic fuzzy partition distance is well-known as an effective measure to determine important attributes. More interestingly, an incremental formula is also developed to quickly compute the intuitionistic fuzzy partition distance in case when the decision table increases in the number of objects. This formula is then applied to construct an incremental attribute reduction algorithm for handling such dynamic tables. Besides, some experiments are conducted on real datasets to show that our method is far superior to the fuzzy rough set based methods in terms of the size of reduct and the classification accuracy.

KEYWORDS

Incremental attribute reduction; intuitionistic fuzzy sets; partition distance measure; dynamic decision tables

1 Introduction

Attribute reduction has long been known as one of the pivotal problems in data preprocessing. The purpose of attribute reduction is to select important attributes and remove redundant or unnecessary ones from decision tables to enhance the efficiency of the classification models, especially for high dimension data. Dubois and Prade applied the fuzzy rough set (FRS) theory to the problem of attribute reduction for addressing directly with the original decision tables consisting of a numerical domain, instead of discretized data. According to this approach, some researches have been published based on fuzzy approximate space. Some typical methods comprise fuzzy membership function [1], fuzzy positive region [2–4], fuzzy information entropy [5–8] and fuzzy distance [9].

In big data trend, the decision tables have often the high dimension and may be updated regularly. The process of adding and removing objects is usually taking place. This problem has brought many difficulties to the traditional attribute reduction approaches. Firstly, the algorithms can meet an obstacle of processing speed and storage space on the high-dimensional decision tables. Secondly, for updated decision tables, the algorithms must also compute the reduct on the entire decision table again which makes to increase the computational time. To deal more effectively with these issues, many researchers proposed some incremental computational techniques for finding the reduct on the dynamic decision table. The incremental methods only update the subset of selected attributes on the altered part of data without re-computing it on the whole decision table. Hence, the processing time can be reduced significantly. Furthermore, with the high-dimensional decision table, it is possible to split data into many parts and apply incremental attribute reduction algorithms. Based on the FRS approach, the incremental algorithms are proposed for solving cases, including addition and removal of the object set [10–14] or the attribute set [15]. Additionally, some researchers extended the incremental algorithms to address incomplete dynamic decision tables. In particular, Giang et al. constructed the tolerance rough set to design hybrid incremental algorithms when supplementing and removing object sets [16]. Afterwards, Thang et al. also developed some formulas for their incremental algorithms for two cases of adding and removing attribute sets [17]. However, Hung et al. showed that the FRS based attribute reduction approach is less efficient in noisy data sets with the low classification accuracy [18].

Recently, some researchers proposed new models using intuitionistic fuzzy sets (IFSs) to solve the attribute selection problem. These models were designed to minimize the noisy information on the decision tables by using a further non-membership function which can adjust noisy objects to give a suitable classification [19]. For such noisy data sets, the IFS based attribute reduction algorithms have often the better processing ability than algorithms using other approaches, such as rough sets or fuzzy rough sets. From the IFSs approach, Tan et al. [20] constructed intuitionistic fuzzy conditional entropy measures and proposed a heuristic algorithm for finding a relative reduct. Then, Thang et al. [21] built a distance measure based on the IFS model. By using the constructed distance measure, they proposed the IFDBAR algorithm in the filter approach to select a subset of important attributes. The experimental results showed that IFDBAR achieved the superior classification ability to the FRS approach on the noisy data sets, especially in case of the low initial classification accuracy.

It can be said that the intuitionistic fuzzy set is a very efficient approach for attribute reduction on the noisy data sets. The subset of features selected from the IFS based algorithms can significantly improve the classification performance of the machine learning models. As mentioned, however, the IFS approach also exists a serious limitation that is the high computational time. Therefore, the attribute reduction algorithms using this approach are often ineffective in dealing with high-dimensional and large datasets. This motivated us to develop a novel incremental method for finding reducts more quickly and efficiently. Specifically, we present an original method to attribute reduction based on intuitionistic fuzzy partition distance. Then we propose an incremental algorithm to extract important attributes from the decision tables in case of the increase of the object number. In our method, an incremental formula is given to fast calculate the intuitionistic fuzzy partition distance on the decision tables when adding an object set. By theoretical and experimental results, we will demonstrate that the proposed method can enhance the performance of the attribute reduction process in comparing with some other state-of-the-art methods in terms of the computational time and classification accuracy.

The paper is organized as follows. The next section will recall several basic concepts of intuitionistic fuzzy sets and related properties. In Section 3, this paper shall define a new reduct based on the intuitionistic fuzzy partition distance and then propose an effective attribute reduction method. This paper will also provide an incremental algorithm for finding reducts on the dynamic decision tables. Section 4 will present experimental results as well as some related analyses. In the last section of this paper, we will draw several conclusions with future works.

2 Preliminary

This section will summarize some basic concepts of intuitionistic fuzzy sets which are an important foundation for proposing an attribute reduction algorithm in the third part of the paper. These basic concepts can be found in [21–23].

First, a decision table is a pair of $DT = (U, C \cup D)$, where U is a finite nonempty set of objects, also known as the universe, C and D are finite nonempty sets of attributes such that each $a \in C \cup D$ determines a map $a\colon U \to V_a$, where V_a is the value set of a. Then, for $u \in U$ and $a \in C \cup D$, the value of a for u is written as $a(u)$. Here, we shall call C as condition attributes and D as decision attributes.

Without losing the comprehensive characteristics, hypothesis D only has one decision attribute d (if D has many attributes, a transformation that can be reduced to an attribute). Accordingly, a decision table can be written as $DT = (U, C \cup \{d\})$.

Given a decision table $DT = (U, C \cup \{d\})$, an intuitionistic fuzzy set (IFS) P on U has the form $P = \{\langle u, \mu_P(u), \vartheta_P(u)\rangle \,|u \in U\}$, in which $\mu_P(u)\colon U \to [0, 1]$ and $\vartheta_P(u)\colon U \to [0, 1]$ are respectively the membership and non-membership degrees of u in P such that $0 \leq \mu_P(u) + \vartheta_P(u) \leq 1, \forall u \in U$ [22].

The hesitant degree of u in P is determined by $\pi_P(u) = 1 - \mu_P(u) - \vartheta_P(u)$. When $\pi_P(u) = 0\ \forall u \in U$, IFS becomes a traditional FS. The cardinality of P is denoted as $|P|$ determined by the formula [23]:

$$|P| = \sum_{u\in U} \frac{1 + \mu_P(u) - \vartheta_P(u)}{2} \tag{1}$$

Consider two IFSs P and Q on U, we will define several set operations to compare them as follows:

1. $P \subseteq Q$ iff $\mu_P(u) \leq \mu_P(u)$ and $\vartheta_P(u) \geq \vartheta_Q(u)$ for any $u \in U$.
2. $P = Q$ iff $P \subseteq Q$ and $Q \subseteq P$.
3. $P \cap Q = \left\{\left(u, min(\mu_P(u), \mu_Q(u)), max\left(\vartheta_P(u), \vartheta_Q(u)\right)\right)\right\}$
4. $P \cup Q = \left\{\left(u, max(\mu_P(u), \mu_Q(u)), min\left(\vartheta_P(u), \vartheta_Q(u)\right)\right)\right\}$.

To facilitate the deployment of definitions and calculation formulas later, an IFS P will sometimes be briefly represented by two components, membership and non-membership. Specifically, $P = \{\mu_P(u), \vartheta_P(u)|u \in U\}$.

Let U be the non-empty finite set of objects. An intuitionistic fuzzy binary relation R on $U \times U$ is defined as follows:

$$R = \{(u, v), \mu_R(u, v), \vartheta_R(u, v) \mid (u, v) \in U \times U\} \quad (2)$$

where $\mu_R(u, v) \in [0, 1]$ and $\vartheta_R(u, v) \in [0, 1]$ are the similarity and diversity functions, respectively. The pair $(\mu_R(u, v), \vartheta_R(u, v))$ is called an intuitionistic fuzzy number between two objects u and v, which satisfies $0 \le \mu_R(u, v) + \vartheta_R(u, v) \le 1$.

Then, R is called an intuitionistic fuzzy equivalence relation (IFER) if R satisfies:

1) Reflexive: iff $\mu_R(u, u) = 1$ and $\vartheta_R(u, u) = 0$ hold for each $u \in U$.
2) Symmetric: iff $\mu_R(u, v) = \mu_R(v, u)$ and $\vartheta_R(u, v) = \vartheta_R(v, u)$ hold for each $u, v \in U$.
3) Transitive: iff $\mu_R(u, v) \ge \sup_{t \in U} \{min(\mu_R(u, t), \mu_R(t, v))\}; \vartheta_R(u, v) \le \inf_{t \in U} \{max(\vartheta_R(u, t), \vartheta_R(t, v))\}$ hold for each $u, v \in U$.

Given an IFER R on U, an attribute subset $A \subseteq C$ and an object $u \in U$, the intuitionistic fuzzy equivalence class (IFEC) of u according to R on A is denoted $R_A[u]$, as follows:

$$R_A[u] = \left\{\left(v, \mu_{R_A[u]}(v), \vartheta_{R_A[u]}(v)\right) \mid v \in U\right\} \quad (3)$$

Consider $DT = (U, C \cup \{d\})$, each $A \subseteq C$ determines an IFER R_A on U. The IFER R_A generates an intuitionistic fuzzy partition (IFP) on U, denoted as ψ_A, $\psi_A = \{R_A[u] \mid u \in U\}$ in which $R_A[u] = \left\{\left(v, \mu_{R_A[u]}(v), \vartheta_{R_A[u]}(v)\right) \mid v \in U\right\}$ is an IFEC of u according to R_A. We can see that each IFEC $R_A[u]$ is also an IFS on U. To simplify the denotation, for each object v, we denote $R_A[u](v) = (\mu_A[u](v), \vartheta_A[u](v))$.

For $A, B \subseteq C$, we have $R_A[u] = \cap_{a \in A} R_{\{a\}}[u]$ and $R_{A \cup B}[u] = R_A[u] \cap R_B[u]$. This means that $R_{A \cup B}[u](v) = (min\{\mu_A[u](v), \mu_B[u](v)\}, max\{\vartheta_A[u](v), \vartheta_B[u](v)\})$ and $\psi_{A \cup B} = \psi_A \cap \psi_B$.

With $A \subseteq C$, there are two special cases:

- If $R_A[u](v) = (0, 1), u \ne v$ and $R_A[u](v) = (1, 0)$ where $v \in U$ then $|R_A[u]| = 1$, the IFP ψ_A is assumed as the smoothest case and it is denoted by ψ_ω.

- If $R_A[u](v) = (1, 0)$ where $v \in U$ then $|R_A[u]| = m$, the IFP ψ_A is assumed as the most unsmooth case and it is denoted by ψ_δ.

Suppose we are given two IFPs ψ_A and ψ_B. We say that ψ_A is finer than ψ_B, denoted $\psi_A \preceq \psi_B$, if for all $u \in U$, $R_A[u] \subseteq R_B[u]$.

3 Attribute Reduction Methods

3.1 Attribute Reduction Based on IFPD

This paper will present in this section a new method for finding a reduct set based on the IFP distance. The cardinal steps of the algorithm are structured as follows. Firstly, we introduce the distance between two IFPs. Then, we define the reduct and the significance of the attribute. Finally, we construct a heuristic algorithm based on the IFP measure.

Given $DT = (U, C \cup \{d\})$ with $U = \{u_1, u_2, \ldots, u_m\}$ and the intuitionistic fuzzy partitions ψ_A, ψ_B generated by the intuitionistic fuzzy equivalence classes $R_A[u_i]$, $R_B[u_i]$, for all $u_i \in U$ and $A, B \subseteq C$, then

$$\mathcal{D}(\psi_A, \psi_B) = \sum_{i=1}^{m} \frac{(|R_A[u_i] \cup R_B[u_i]| - |R_A[u_i] \cap R_B[u_i]|)}{m^2} \tag{4}$$

is an intuitionistic fuzzy partition distance (IFPD) between ψ_A and ψ_B [21].

It is easy to see that the minimum value of $\mathcal{D}(\psi_A, \psi_B)$ is equal to 0 when $\psi_A = \psi_B$ and the maximum value of $\mathcal{D}(\psi_A, \psi_B)$ reaches as $\dfrac{m-1}{m}$ if and only if $\psi_A = \psi_\delta$ and $\psi_B = \psi_\omega$ (or $\psi_A = \psi_\omega$ and $\psi_B = \psi_\delta$). Thus, we have $0 \le \mathcal{D}(\psi_A, \psi_B) \le \dfrac{m-1}{m}$.

Based on $\mathcal{D}(\psi_A, \psi_B)$, IFPD created by C and $C \cup \{d\}$ on U is computed by:

$$\mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \sum_{i=1}^{m} \frac{\left(|R_C[u_i]| - \left|R_C[u_i] \cap R_{\{d\}}[u_i]\right|\right)}{m^2} \tag{5}$$

Clearly, if $B \subseteq C$, then $\mathcal{D}\left(\psi_B, \psi_{B\cup\{d\}}\right) \ge \mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right)$. Hence, the intuitionistic fuzzy partition distance satisfies the non-monotonous property with respect to the set of condition attributes. It implies that the size of B is the inverse of the value of $\mathcal{D}\left(\psi_B, \psi_{B\cup\{d\}}\right)$. Then, the greater the value of $\mathcal{D}\left(\psi_B, \psi_{B\cup\{d\}}\right)$ is, the smaller the size of B is. We can use the intuitionistic fuzzy partition distance as a criterion for choosing attributes in the process of attribute reduction.

Definition 1. Let $DT = (U, C \cup \{d\})$. A subset $M \subseteq C$ is called a reduct of C if

1. $\mathcal{D}\left(\psi_M, \psi_{M\cup\{d\}}\right) = \mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right)$,
2. $\forall M' \subset M, \mathcal{D}\left(\psi_{M'}, \psi_{M'\cup\{d\}}\right) > \mathcal{D}\left(\psi_M, \psi_{M\cup\{d\}}\right)$.

This definition implies that for any $b \in M$, if $\mathcal{D}\left(\psi_{M\setminus\{b\}}, \psi_{M\setminus\{b\}\cup\{d\}}\right) \neq \mathcal{D}\left(\mathcal{P}_C, \mathcal{P}_{C\cup\{d\}}\right)$, then b is indispensable in M. In contrast, b will be called a redundant attribute in M.

Definition 2. Let $DT = (U, C \cup \{d\})$, $M \subseteq C$ and $b \in C/M$. Then the significance measure of b with respect to M, denoted $SIG_M(b)$, is determined by the formula:

$$SIG_M(b) = \mathcal{D}\left(\psi_M, \psi_{M\cup\{d\}}\right) - \mathcal{D}\left(\psi_{M\cup\{b\}}, \psi_{M\cup\{b\}\cup\{d\}}\right) \tag{6}$$

It can be easily seen that $SIG_M(b) \ge 0$. Consider any attribute $b \in C$, its significance for an attribute subset characterizes the classification quality of b. We can see the alteration of the certainty degree. If the value of $SIG_M(b)$ is higher, then the attribute b will be more essential. This measure can be considered as a criterion for selecting the necessary attributes. Based on this definition, we design an effective algorithm to extract an optimal attribute subset from a given decision table. The algorithm begins with an empty set and supplements one attribute with the highest significance into the selected feature subset at each iteration until the condition stops happening. In particular, our attribute reduction method, denoted as Algorithm 1, is designed as illustrated below, and its flowchart can be found in Fig. 1.

Algorithm 1: Attribute reduction based on the IFPD (ARIFPD)

Input: Decision Table $DT = (U, C \cup \{d\})$
Output: Approximation reduct M
1. $M: = \varnothing, \mathcal{D}\left(\psi_M, \psi_{M\cup\{d\}}\right) = 1$
2. Calculate $\mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right)$ using (5)
3. **While** $\mathcal{D}\left(\psi_M, \psi_{M\cup\{d\}}\right) > \mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right)$ **do:**
4. **For** each $b \in C\backslash M$ **do:**
5. Compute $\mathcal{D}\left(\psi_{M\cup\{b\}}, \psi_{M\cup\{b\}\cup\{d\}}\right)$
6. **End**
7. Select b_0 which satisfies $SIG_M(b_0) = \max\limits_{b\in C\backslash M}\{SIG_M(b)\}$
8. $M: = M \cup \{b_0\}$
9. **End**
10. Return M

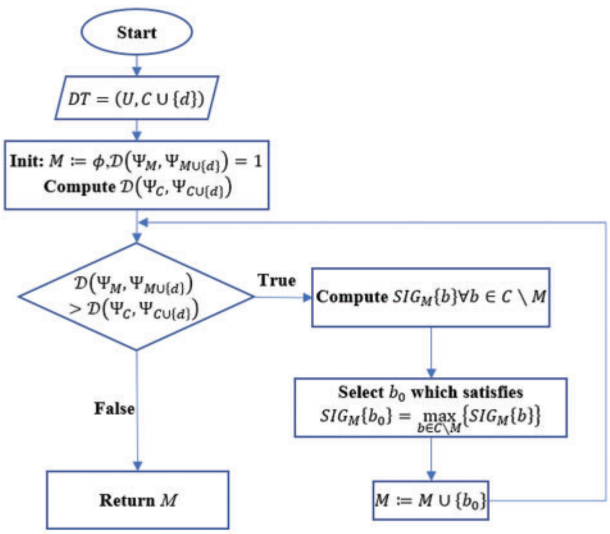

Figure 1: Flowchart of Algorithm 1

We now examine the computational complexity of ARIFPD. Suppose that $|C|, |U|$ are the number of condition attributes and instances on the decision table, respectively. It is clear that IFP ψ_C can be determined in $\mathcal{O}\left(|C| * |U|^2\right)$ time. Hence, the computational complexity of calculating $\mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right)$ in line 2 is $\mathcal{O}\left(|C| * |U|^2\right)$. Besides, it can be easily seen that the computational cost of $\mathcal{D}\left(\psi_{M\cup\{b\}}, \psi_{M\cup\{b\}\cup\{d\}}\right)$ is $\mathcal{O}\left(|U|^2\right)$, and $SIG_M(b)$ in line 7 is calculated in $\mathcal{O}\left(|U|^2\right)$ time. Thus, the computational time of the For loop from line 4 to line 6 is $\mathcal{O}\left(|C| * |U|^2\right)$ and the computational time of the loop While from line 3 to line 9 is $\mathcal{O}\left(|C|^2 * |U|^2\right)$. Summary, ARIFPD has the time complexity to be $\mathcal{O}\left(|C|^2 * |U|^2\right)$.

3.2 Incremental Attribute Reduction Algorithm

We will propose this section an incremental algorithm using IFPD when supplementing a new object set into a given decision table. We now begin with providing a formula to quickly calculate IFPD after supplementing an object set which is extended in [21].

Proposition 1. Given $DT = (U, C \cup \{d\})$ with $U = \{u_1, u_2, \ldots, u_m\}$, intuitionistic fuzzy equivalence relations R_C, $R_{\{d\}}$ and a new object set $\Delta U = \{u_{m+1}, u_{m+2}, \ldots, u_{m+s}\}$, $s \geq 1$, the IFPD between ψ_C and $\psi_{C\cup\{d\}}$ on $U \cup \Delta U$ is determined by:

$$\mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \frac{m^2 * \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right)}{(m+s)^2} + 2 * \sum_{i=1}^{s} \frac{\left(\left|R_C\left[u_{m+i}\right]\right| - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right| - \zeta_i\right)}{(m+s)^2} \quad (7)$$

where $\zeta_1 = 0$ and for $i \geq 2$, $\zeta_i = \frac{1}{2}\sum_{j=i}^{s-1}\begin{pmatrix}\mu_C\left[u_{m+i}\right]\left(u_{m+j+1}\right) - \min\left\{\mu_C\left[u_{m+i}\right]\left(u_{m+j+1}\right), \mu_{\{d\}}\left[u_{m+i}\right]\left(u_{m+j+1}\right)\right\} \\ -\vartheta_C\left[u_{m+i}\right]\left(u_{m+j+1}\right) + \max\left\{\vartheta_C\left[u_{m+i}\right]\left(u_{m+j+1}\right), \vartheta_{\{d\}}\left[u_{m+j}\right]\left(u_{m+j+1}\right)\right\}\end{pmatrix}$

Proof:

It is easy to see that if $s = 1$, we have:

$$\mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \frac{m^2 * \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right)}{(m+s)^2} + 2 * \sum_{i=1}^{s} \frac{\left(\left|R_C\left[u_{m+i}\right]\right| - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right|\right)}{(m+s)^2}$$

To simplify for the proof of this proposition, we will set:

$$\zeta_{ij} = \frac{1}{2}\left(\mu_C\left[u_i\right]\left(u_j\right) - \min\left\{\mu_C\left[u_i\right]\left(u_j\right), \mu_{\{d\}}\left[u_i\right]\left(u_j\right)\right\} - \vartheta_C\left[u_i\right]\left(u_j\right) + \max\left\{\vartheta_C\left[u_i\right]\left(u_j\right), \vartheta_{\{d\}}\left[u_i\right]\left(u_j\right)\right\}\right)$$

Since U is added s objects, the formula in (5) becomes:

$$\mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \frac{\sum_{i=1}^{m+s}\left(\left|R_C\left[u_i\right]\right| - \left|R_C\left[u_i\right] \cap R_{\{d\}}\left[u_i\right]\right|\right)}{(m+s)^2}$$

$$= \frac{1}{(m+s)^2}\left(\sum_{i=1}^{m+s}\left(\zeta_{1i}\right) + \ldots + \sum_{i=1}^{m+s}\left(\zeta_{mi}\right) + \sum_{i=1}^{s}\left(\left|R_C\left[u_{m+i}\right]\right| - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right|\right)\right)$$

$$= \frac{1}{(m+s)^2}\begin{pmatrix}\sum_{i=1}^{m}\left(\zeta_{1i}\right) + \ldots + \sum_{i=1}^{m}\left(\zeta_{mi}\right) + \sum_{i=1}^{s}\left(\zeta_{1m+i}\right) + \ldots + \sum_{i=1}^{s}\left(\zeta_{mm+i}\right) + \sum_{i=1}^{s}\left(\zeta_{m+1m+i}\right) + \ldots \\ + \sum_{i=1}^{s}\left(\zeta_{m+sm+i}\right) - \left(\sum_{i=1}^{s}\left(\zeta_{m+1m+i}\right) + \ldots + \sum_{i=1}^{s}\left(\zeta_{m+sm+i}\right)\right) + \sum_{i=1}^{s}\left(\left|R_C\left[u_{m+i}\right]\right|\right. \\ \left. - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right|\right)\end{pmatrix}$$

We have

$$\sum_{i=1}^{s}\left(\zeta_{1m+i}\right) + \ldots + \sum_{i=1}^{s}\left(\zeta_{mm+i}\right) + \sum_{i=1}^{s}\left(\zeta_{m+1m+i}\right) + \ldots + \sum_{i=1}^{s}\left(\zeta_{m+sm+i}\right) = \sum_{i=1}^{s}\left(\left|R_C\left[u_{m+i}\right]\right| - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right|\right),$$

$$\sum_{i=1}^{m}\left(\zeta_{1i}\right) + \ldots + \sum_{i=1}^{m}\left(\zeta_{mi}\right) = m^2 * \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right), \left(\sum_{i=1}^{s}\left(\zeta_{m+1m+i}\right) + \ldots + \sum_{i=1}^{s}\left(\zeta_{m+sm+i}\right)\right) = \sum_{i=1}^{s}\sum_{j=1}^{s}\left(\zeta_{m+im+j}\right),$$

thus $\mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \frac{1}{(m+s)^2}\begin{pmatrix}m^2 * \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right) + 2 * \sum_{i=1}^{s}\left(\left|R_C\left[u_{m+i}\right]\right| - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right|\right) \\ -\sum_{i=1}^{s}\sum_{j=1}^{s}\left(\zeta_{m+im+j}\right)\end{pmatrix}$

Furthermore, $\zeta_{ii} = 0$ and $\zeta_{ij} = \zeta_{ji}$, we get $\sum_{i=1}^{s}\sum_{j=1}^{s}\left(\zeta_{m+im+j}\right) = 2\sum_{i=1}^{s}\sum_{j=i}^{s-1}\left(\zeta_{m+im+j+1}\right) = 2\sum_{i=1}^{s}\zeta_i$.

Hence, $\mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \frac{m^2 * \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right)}{(m+s)^2} + 2 * \sum_{i=1}^{s}\frac{\left(\left|R_C\left[u_{m+i}\right]\right| - \left|R_C\left[u_{m+i}\right] \cap R_{\{d\}}\left[u_{m+i}\right]\right| - \zeta_i\right)}{(m+s)^2}$.

The formula (6) consists of two main components. The first component computes the distance of two intuitionistic fuzzy partitions without adding the set of new objects. This component is already determined from the previous stages. Therefore, the incremental formula provided in Proposition 1 allows us to only executes on the second component consisting of the intuitionistic fuzzy equivalence classes generated by the additions of the new object set. From this formula, we will replace the formula (5) to conduct on the decision table when supplementing new objects. Accordingly, the algorithm will reduce the computational complexity and obtain a proximate reduct set.

Proposition 2. Let $DT = (U, C \cup \{d\})$ with $U = \{u_1, u_2, \ldots, u_m\}$. Suppose that $M \subseteq C$ is a reduct based on IFPD on U and $\Delta U = \{u_{m+1}, u_{m+2}, \ldots, u_{m+s}\}$, $s \geq 1$, is an incremental object set. Then

1. If all objects in ΔU have the same value of decision attribute then

$$\mathcal{D}_{U \cup \Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \frac{m^2 * \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right)}{(m+s)^2} + 2 * \sum_{i=1}^{s} \frac{\left(|R_C[u_{m+i}]| - \left|R_C[u_{m+i}] \cap R_{\{d\}}[u_{m+i}]\right|\right)}{(m+s)^2}$$

2. If $R_M[u_{m+i}] \subseteq R_{\{d\}}[u_{m+i}]$ for $i = 1, 2, \ldots, s$ then $\mathcal{D}_{U\cup\Delta U}\left(\psi_M, \psi_{M\cup\{d\}}\right) = \mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right)$

Proof:

For any $i = 1, 2, \ldots, s$ and $j = 1, 2, \ldots, (s-1)$, we consider:

1. If all objects in ΔU have the same value of decision attribute, then we have $\mu_{\{d\}}[u_{m+i}]\left(u_{m+j+1}\right) = 1$ and $\vartheta_{\{d\}}\left[u_{m+j}\right]\left(u_{m+j+1}\right) = 0$. Thus, $\min\left\{\mu_C[u_{m+i}]\left(u_{m+j+1}\right), \mu_{\{d\}}[u_{m+i}]\left(u_{m+j+1}\right)\right\} = \mu_C[u_{m+i}]\left(u_{m+j+1}\right)$ and $\max\left\{\vartheta_C[u_{m+i}]\left(u_{m+j+1}\right), \vartheta_{\{d\}}\left[u_{m+j}\right]\left(u_{m+j+1}\right)\right\} = \vartheta_C[u_{m+i}]\left(u_{m+j+1}\right)$, from Proposition 1, the formula in the first case is obtained.
2. If $R_M[u_{m+i}] \subseteq R_{\{d\}}[u_{m+i}]$ then $R_C[u_{m+i}] \subseteq R_M[u_{m+i}] \subseteq R_{\{d\}}[u_{m+i}]$.

Hence $\left|R_M[u_{m+i}] \cap R_{\{d\}}[u_{m+i}]\right| = |R_M[u_{m+i}]|$ and $\left|R_C[u_{m+i}] \cap R_{\{d\}}[u_{m+i}]\right| = |R_C[u_{m+i}]|$. Thus, we get these equations $|R_M[u_{m+i}]| - \left|R_M[u_{m+i}] \cap R_{\{d\}}[u_{m+i}]\right| = 0$ and $|R_C[u_{m+i}]| - \left|R_C[u_{m+i}] \cap R_{\{d\}}[u_{m+i}]\right| = 0$. More importantly,

$$\begin{aligned}
&\min\left\{\mu_M[u_{m+i}]\left(u_{m+j+1}\right), \mu_{\{d\}}[u_{m+i}]\left(u_{m+j+1}\right)\right\} = \mu_M[u_{m+i}]\left(u_{m+j+1}\right),\\
&\max\left\{\vartheta_M[u_{m+i}]\left(u_{m+j+1}\right), \vartheta_{\{d\}}\left[u_{m+j}\right]\left(u_{m+j+1}\right)\right\} = \vartheta_M[u_{m+i}]\left(u_{m+j+1}\right),\\
&\min\left\{\mu_C[u_{m+i}]\left(u_{m+j+1}\right), \mu_{\{d\}}[u_{m+i}]\left(u_{m+j+1}\right)\right\} = \mu_C[u_{m+i}]\left(u_{m+j+1}\right),\\
&\max\left\{\vartheta_C[u_{m+i}]\left(u_{m+j+1}\right), \vartheta_{\{d\}}\left[u_{m+j}\right]\left(u_{m+j+1}\right)\right\} = \vartheta_C[u_{m+i}]\left(u_{m+j+1}\right).
\end{aligned}$$

Based on Proposition 1, we get:

$$\mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right) = \left(\frac{m}{m+s}\right)^2 \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right) \text{ and } \mathcal{D}_{U\cup\Delta U}\left(\psi_M, \psi_{M\cup\{d\}}\right) = \left(\frac{m}{m+s}\right)^2 \mathcal{D}_U\left(\psi_M, \psi_{M\cup\{d\}}\right).$$

Because M is a reduct of C, from the definition 1, we obtain $\mathcal{D}_U\left(\psi_M, \psi_{M\cup\{d\}}\right) = \mathcal{D}_U\left(\psi_C, \psi_{C\cup\{d\}}\right)$. Thus $\mathcal{D}_{U\cup\Delta U}\left(\psi_M, \psi_{M\cup\{d\}}\right) = \mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right)$. The proof is complete.

The result of Proposition 2 shown that the incremental algorithm for IFPD based Attribute Reduction when Adding Objects (ARIFPD_AO) comprises four main steps as in Algorithm 2 and its corresponding flowchart is visually represented in Fig. 2.

Algorithm 2: Algorithm (ARIFPD_AO)

Input:

1. Decision Table $DT = (U, C \cup \{d\})$ with $U = \{u_1, u_2, \ldots, u_m\}$, an IFER R, the reduct $M \subseteq C$.
2. Intuitionistic fuzzy partitions ψ_M and ψ_C
3. Added set of objects $\Delta U = \{u_{m+1}, u_{m+2}, \ldots, u_{m+s}\}$

Output: The approximation reduct $M_{U\cup\Delta U}$ of $DT' = (U \cup \Delta U, C \cup \{d\})$

// Initialization

1. $M_{U\cup\Delta U} := M$
2. Compute intuitionistic fuzzy partitions on $U \cup \Delta U$: $\psi_M, \psi_{\{d\}}$

// Check the added set of objects

3. Set $S: = \Delta U$
4. **For** i = 1 to s **do:**
5.   **If** $R_M[u_{m+i}] \subseteq R_{\{d\}}[u_{m+i}]$ **then** $S := S \setminus \{u_{m+i}\}$
6. **If** $S = \varnothing$ then **return** $M_{U\cup\Delta U}$ // Approximation reduct does not change
7. Set $\Delta U: = S, s: = |\Delta U|$ //Update the incremental object set.

// Finding the reduct

8. Compute original IFPDs: $\mathcal{D}\left(\psi_{M_{U\cup\Delta U}}, \psi_{M_{U\cup\Delta U}\cup\{d\}}\right), \mathcal{D}\left(\psi_C, \psi_{C\cup\{d\}}\right)$
9. Compute IFPDs using the incremental formula: $\mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}}, \psi_{M_{U\cup\Delta U}\cup\{d\}}\right), \mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right)$
10. While $\mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}}, \psi_{M_{U\cup\Delta U}\cup\{d\}}\right) > \mathcal{D}_{U\cup\Delta U}\left(\psi_C, \psi_{C\cup\{d\}}\right)$ **do:**
11.   **For** each $b \in C \setminus M_{U\cup\Delta U}$ **do:**
12.   **Compute** $\mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}\cup\{b\}}, \psi_{M_{U\cup\Delta U}\cup\{b\}\cup\{d\}}\right)$ by using the incremental Formula (7)
13.   **Compute** $SIG_{M_{U\cup\Delta U}}(b) = \mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}}, \psi_{M_{U\cup\Delta U}\cup\{d\}}\right) - \mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}\cup\{b\}}, \psi_{M_{U\cup\Delta U}\cup\{b\}\cup\{d\}}\right)$
14.   **End**
15.   Select $b_0$which satisfies: $SIG_{M_{U\cup\Delta U}}(b_0) = \underset{b\in C\setminus M_{U\cup\Delta U}}{Max}\left\{SIG_{M_{U\cup\Delta U}}(b)\right\}$
16. $M_{U\cup\Delta U} := M_{U\cup\Delta U} \cup \{b_0\}$

// Remove redundant attributes

17. For each $b \in M_{U\cup\Delta U}$ **do:**
18.   **Compute** $\mathcal{D}_{U\cup\Delta U}\left(\psi_{B\setminus\{b\}}, \psi_{M_{U\cup\Delta U}\setminus\{b\}\cup\{d\}}\right)$ by using the incremental Formula (7)
19.   **If** $\mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}\setminus\{b\}}, \psi_{M_{U\cup\Delta U}\setminus\{b\}\cup\{d\}}\right) = \mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}}, \psi_{M_{U\cup\Delta U}\cup\{d\}}\right)$ **then** $M_{U\cup\Delta U} := M_{U\cup\Delta U} \setminus \{b\}$
20. End
21. Return $M_{U\cup\Delta U}$

We continue to evaluate the computational complexity of the algorithm ARIFPD_AO. We use 3 symbols $|C|, |U|$ and $|\Delta U|$ to denote the number of condition attributes, original instances, and instances supplemented to the original set, respectively. The computational complexity of ARIFPD_AO is calculated by the pseudocode above. The computational complexity of computing IFP in line 2 is $\mathcal{O}(|M_{U\cup\Delta U}| * |\Delta U| * (|U| + |\Delta U|))$. In the simplest case, the algorithm stops in line 6. Then, the computational time of ARIFPD_AO is $\mathcal{O}(|M_{U\cup\Delta U}| * |\Delta U| * (|U| + |\Delta U|))$. For the remaining case the computational time IFPDs in line 9 is $\mathcal{O}(|C| * |\Delta U| * (|U| + |\Delta U|))$, and the incremental calculation $\mathcal{D}_{U\cup\Delta U}\left(\psi_{M_{U\cup\Delta U}\cup\{b\}}, \psi_{M_{U\cup\Delta U}\cup\{b\}\cup\{d\}}\right)$ has a time complexity of $\mathcal{O}(|\Delta U| * (|U| + |\Delta U|))$. Similar to the computational complexity of ARIFPD, the complexity of the loop While in algorithm ARIFPD_AO is $\mathcal{O}\left((|C| - |M_{U\cup\Delta U}|)^2 * |\Delta U| * (|U| + |\Delta U|)\right)$. From lines 17 to 20, the time complexity of the loop For is $\mathcal{O}\left(|C|^2 * |\Delta U| * (|U| + |\Delta U|)\right)$. Therefore, the complexity of the algorithm in the worst case is $\mathcal{O}\left(|C|^2 * |\Delta U| * (|U| + |\Delta U|)\right)$. The computational complexity of the algorithm is $max\left\{\mathcal{O}\left(|M_{U\cup\Delta U}| * |\Delta U| * (|U| + |\Delta U|)\right), \mathcal{O}\left(|C|^2 * |\Delta U| * (|U| + |\Delta U|)\right)\right\}$. Thus, the incremental algorithm ARIFPD_AO has the ability to reduce the time complexity.

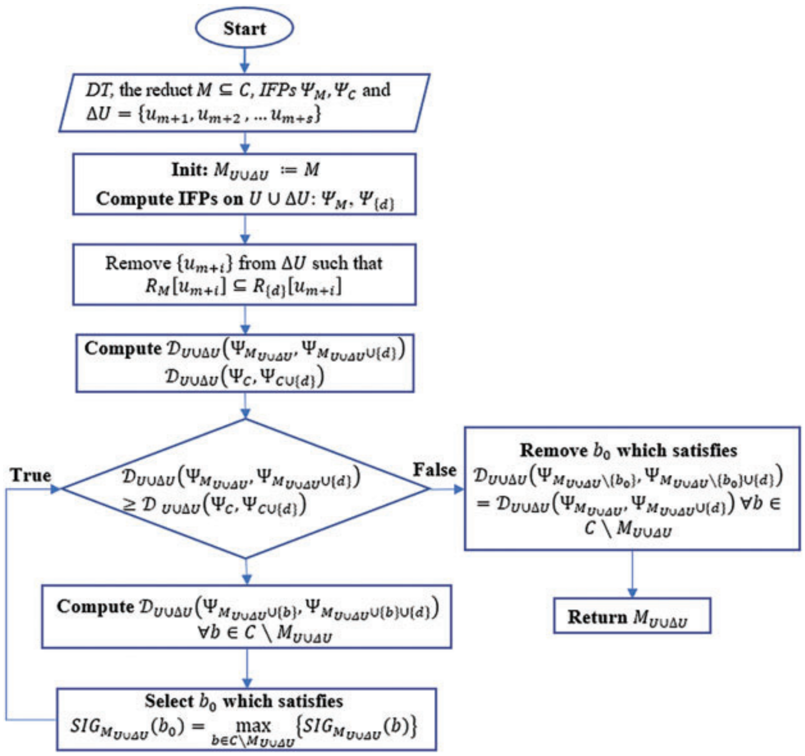

Figure 2: Flowchart of Algorithm 2

4 Experiments

In the previous sections, this paper presented an incremental approach based on IFPD to process the dynamic decision table. The paper will demonstrate this part with some experiments to prove the efficiency of our method (ARIFPD_AO). The paper compares the efficiency of the ARIFPD_AO, IFSA in [24] and FDAR_AO in [10] through three criteria: Classification accuracy performance based on classifier KNN (k =10) with tenfold cross-validation, size of reduct and computation time.

4.1 Experiments Setup

Experimental data: To prove the efficiency of the proposed method, we conduct experiments on some benchmark datasets from the UCI Machine learning repository [25]. The experimental data focuses mainly on sets with low initial classification accuracy and with a large number of instances. Data set is split up two subsets with the same number of objects. The first subset is denoted by U_{ori} (Column 5 in Table 1) is used in the algorithm ARIFPD to find the reduct and U_{inc} (Column 6 in Table 1) is used in the incremental algorithm ARIFPD_AO. Next, the incremental set U_{inc} is separated into six equal parts U_1, U_2, U_3, U_4, U_5, U_6 respectively.

In Table 1, columns $|U|$, $|U_{ori}|$, $|U_{inc}|$, $|C|$, $|d|$ are used to denote for the number of objects in each data set, objects in U_{ori}, objects in U_{inc}, conditional attributes, decision classes, respectively.

Table 1: The description of data sets used in experiments

ID	Data sets	Description	$\lvert U \rvert$	$\lvert U_{ori} \rvert$	$\lvert U_{inc} \rvert$	$\lvert C \rvert$	$\lvert d \rvert$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Robot-failures	Robot Execution Failures	164	82	82	90	5
2	Madelon	Madelon datasets	780	390	390	500	2
3	Ionosphere	Ionosphere dataset	351	175	176	34	2
4	Pc4	Data from flight software	1458	729	729	37	2
5	Spectf	SPECTF heart data	349	174	175	44	2
6	Movement-libras	LIBRAS movement	360	180	180	90	15
7	Ozone	Ozone level detection	2534	1267	1267	72	2
8	ORL	Database of face	400	200	200	1024	40
9	Mfeat	Mfeat fourier	2000	1000	1000	76	10
10	Wall-robot	Wall robot navigation	5456	2728	2728	24	4
11	Hill-valley	Hill valley dataset	1212	606	606	100	2

Experimental scenario:

First, we use algorithms ARIFPD, GFS [25] and FDAR [10] to find the reduct on U_{ori}. Next, based on the reducts obtained from the three algorithms, we evaluate the incremental algorithms ARIFPD_AO, IFSA and FDAR_AO from U_1 to U_6 of U_{inc}.

Now, we construct IFERs including the IF similarity degree and IF diversity degree between two objects u and v with respect to the attribute a.

- If the value domain of a is a continuous value type, then:

$$\mu_{\{a\}}[u](v) = 1 - \frac{a(u) - a(v)}{\max(a) - \min(a)} \tag{8}$$

The above formula determines the intuitionistic fuzzy similarity degree of the object u with the object v, in which min(a) and max(a) are minimal and maximal values corresponding to a. In essence, the denominator component of the formula above is the process of min-max data normalizing to ensure that the values in the decision table are in the range [0,1]. Finally, we compute the intuitionistic fuzzy diversity degree based on the below formula.

$$\vartheta_{\{a\}}[u](v) = \frac{1 - \mu_{\{a\}}[u](v)}{1 + \lambda_a * \mu_{\{a\}}[u](v)} \; with \; \lambda_a > 0 \tag{9}$$

Clearly, if $\lambda_a = 0$ then $\upsilon_{\{a\}}[u](v) = 1 - \mu_{\{a\}}[u](v)$ and the pairs $\upsilon_{\{a\}}[u](v)$ and $\upsilon_{\{a\}}[u](v)$ will only present the characteristics of FRS. With $\lambda_a > 0$, we can see that the intuitionistic fuzzy similarity and diversity degrees are inversely proportional to each other and satisfy $0 \le \mu[u](v) + \upsilon[u](v) \le 1$. We recommend the value λ_a of the attribute a according to the following formula:

$$\lambda_a = \begin{cases} 1 \; if \quad \sigma_a = 0 \\ \dfrac{\beta_a}{\sigma_a} \; if \quad \sigma_a > 0 \end{cases} \tag{10}$$

in which $\sigma_a = \sqrt{\frac{1}{|U|-1}\sum_{u\in U}(a(u)-\overline{a})^2}$ is the standard deviation of the value domain of the attribute a, $\beta_a = \frac{\left|\mathcal{P}^F_{\{a\}\cup\{d\}}\right|}{\left|\mathcal{P}^F_{\{d\}}\right|}$ is the consistency of the attribute a in the decision table, $\mathcal{P}^F_{\{a\}\cup\{d\}}$ is the fuzzy partition of $\{a\}\cup\{d\}$ and $\mathcal{P}^F_{\{d\}}$ is the fuzzy partition of d. Clearly, if $\mu_{\{a\}}[u](v)$ has a small value, it will lead to a small consistency of a and $\vartheta_{\{a\}}[u](v)$ has a big value.

- If the value of a is a categorical value, then:

$$\mu_A[u](v) = \begin{cases} 1, a(u) = a(v) \\ 0, a(u) \neq a(v) \end{cases} \tag{11}$$

and

$$\vartheta_A[u](v) = 1 - \mu_A[u](v) \tag{12}$$

4.2 Experimental Results

As mentioned above, the paper first compares the efficiency of the FDAR, GFS and ARIFPD algorithms. The size and the classification accuracy of reducts are shown in Table 2. Across all the data sets, it is clear that the reducts obtained from ARIFPD are often the smallest size, while the reducts obtained by FDAR still have large sizes on some datasets. The paper next compares the computational time between the three algorithms. The time of the algorithms is calculated after the step data preprocessing to when the reduct of the algorithms is determined. The results from Table 2 show shorter times when running the GFS algorithm on whole datasets. The reason for this is because the algorithms based on FRSs and IFSs must compute relational matrices with many elements. Besides, the FDAR algorithm only uses the similarity degree to calculate, while the ARIFPD algorithm has to calculate both the similarity and the diversity degree. Thus, the computational time of ARIFPD is the most complex. Next, this paper compares three algorithms through the KNN classifier to evaluate the classification capacity of reducts. Table 2 and Fig. 3 display the comparative results, in which the raw data column is the classification accuracy when we use the whole attributes of each data set to appreciate, and columns according to three methods provide the classification accuracies appreciated through the attribute subset chose by those algorithms. It can be emphasized that our method determines the important attributes very efficiently for different data sets. More especially, the comparison with raw data shows that the classification performance of the reduct from the proposed algorithm is superiority over the raw data in 9 cases. There is only one case where our reducts have lower classification accuracy than the raw data. In addition, the average classification accuracy of the method in this study reached 76.9% and the raw data reached 73.1%. Therefore, it indicated that the classification accuracy of the proposed algorithm is significantly higher than the original data. Even more intriguingly, ARIFPD's reducts yield superior classification results compared to FDAR and GFS across nearly all datasets, despite ARIFPD selecting a smaller number of attributes, as evidenced in Figs. 3 and 4.

It can be observed more clearly from Table 2 and Fig. 3 that there is one case in the Movement dataset where our reduct has the lower accuracy. However, the average accuracy of our algorithm is also higher than the FDAR and GFS algorithms. It showed that the proposed algorithm has the capacity to select significant attributes for better improvement on the noisy data sets with low classification accuracy.

Table 2: The process results of the FDAR, GFS and ARIFPD

ID	Data sets	RAW	FDAR			GFS			ARIFPD		
		Acc	|B|	Acc	Time	|B|	Acc	Time	|B|	Acc	Time
1	Robot-failures	0.583	8	0.657	0.365	6	0.669	0.510	**2**	**0.682**	0.386
2	Madelon	0.549	41	0.536	63.43	12	0.479	39.15	**11**	**0.615**	74.68
3	Ionosphere	0.771	11	0.806	0.582	11	0.789	0.440	**7**	**0.812**	0.625
4	Pc4	0.883	7	0.875	9.106	8	0.879	6.112	**2**	**0.889**	10.72
5	Spectf	0.766	11	**0.787**	0.549	5	0.770	0.460	**2**	0.781	0.540
6	Movement	0.811	28	**0.817**	1.810	9	0.789	1.885	**8**	0.794	1.936
7	Ozone	0.909	12	0.910	58.94	6	**0.912**	20.90	**2**	0.909	63.42
8	ORL	0.725	142	0.765	62.58	78	0.760	41.50	**66**	**0.805**	73.81
9	Mfeat	0.906	45	0.924	53.86	12	0.924	17.43	**11**	**0.937**	59.05
10	Wall-robot	0.636	17	0.602	96.43	18	0.610	40.75	**4**	**0.714**	109.3
11	Hill-valley	0.510	3	0.502	18.32	12	0.492	11.52	**2**	**0.521**	19.88
AVERAGE		0.731	29.5	0.743	**33.27**	16.1	0.734	16.42	**10.6**	**0.769**	37.66

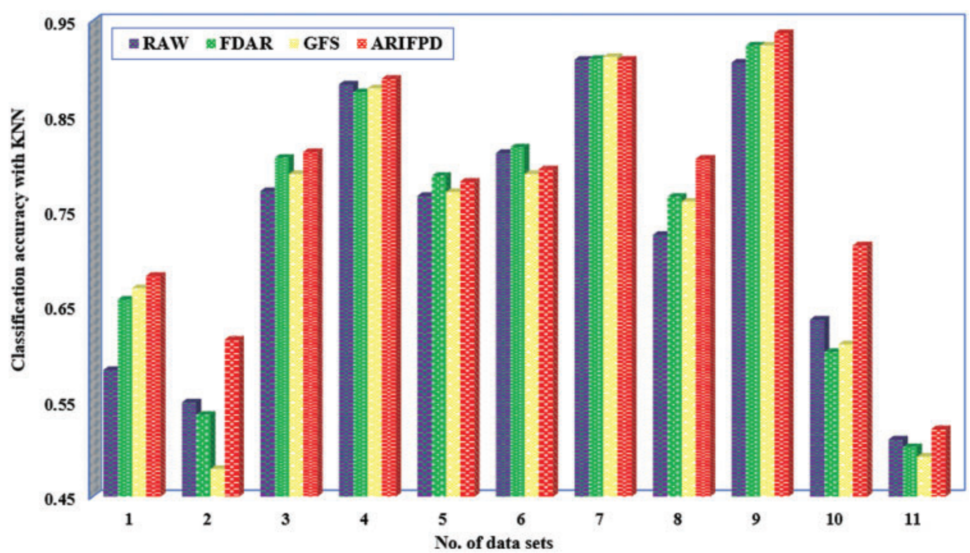

Figure 3: The classification accuracy of the FDAR, GFS and ARIFPD

The performance of the proposed incremental method is appreciated by comparing the FDAR_AO, IFSA and ARIFPD_AO algorithms. First, it is obvious that the incremental algorithms have much faster processing time than the FDAR, IFSA and ARIFPD algorithms because the incremental algorithms compute on the additional parts of the data tables, instead of the whole data table.

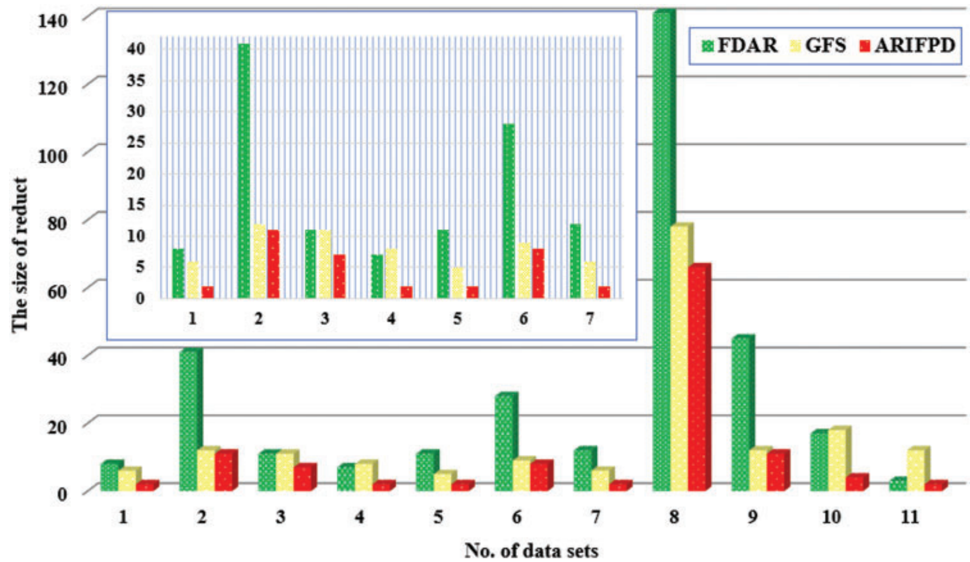

Figure 4: The size of reduct of the FDAR, GFS and ARIFPD

Table 3 shows that for most datasets, the execution time of FDAR_AO and IFSA is faster when compared to ARIFPD_AO. It can be also explained similarly with comparing the execution times of the algorithms FDAR and ARIFPD. Furthermore, the proposed algorithm includes a processing step to remove redundant attributes. Hence the processing time of our algorithm will be slower. However, the execution time of ARIFPD_AO is better than FDAR_AO on some data sets, for example Robot-Failure and Ozone. This is because our obtained reduct is smaller than two remaining algorithms. Then the number of loops is conducted less. The accuracy and size of reducts extracted by our method are also investigated in Table 3. For the size of reduct at each incremental phase, ARIFD_AO is much smaller than FDAR_AO and IFSA, especially for several data sets with a large number of attributes, such as Robot-Failure, Libras-Movement.

Table 3: The process results of the FDAR_AO, IFSA and ARIFPD_AO

ID	Data sets	Adding data sets	RAW	FDAR_AO			IFSA			ARIFPD_AO		
			Acc	$\|B\|$	Acc	Time	$\|B\|$	Acc	Time	$\|B\|$	Acc	Time
1	Robot-failures	$\|U_1\| = 95$	0.610	12	0.630	0.060	13	0.598	4.671	**2**	**0.703**	0.001
		$\|U_2\| = 108$	0.575	13	0.585	0.011	13	0.557	0.005	**3**	**0.631**	0.022
		$\|U_3\| = 121$	0.512	18	0.553	0.104	13	0.504	0.006	**5**	**0.611**	0.077
		$\|U_4\| = 134$	0.483	33	0.529	0.303	13	0.485	0.210	**7**	**0.552**	0.102
		$\|U_5\| = 147$	0.477	33	0.524	0.001	14	0.504	0.008	**8**	**0.608**	0.030
		$\|U_6\| = 164$	0.505	33	0.549	0.001	14	0.529	0.010	**8**	**0.652**	0.001
2	Madelon	$\|U_1\| = 455$	0.580	43	0.560	2.629	14	0.519	1.594	**11**	**0.596**	0.004
		$\|U_2\| = 520$	0.581	44	0.544	2.045	15	0.523	0.485	**11**	**0.594**	0.005
		$\|U_3\| = 585$	0.593	45	0.561	1.894	15	0.511	0.607	**14**	**0.607**	28.17
		$\|U_4\| = 650$	0.609	46	0.546	2.322	16	0.518	0.005	**14**	**0.602**	0.009
		$\|U_5\| = 715$	0.611	47	0.534	2.855	16	0.508	0.722	**14**	**0.617**	0.015
		$\|U_6\| = 780$	0.603	47	0.544	0.005	16	0.496	0.007	**14**	**0.614**	0.012

(Continued)

Table 3 (continued)

ID	Data sets	Adding data sets	RAW	FDAR_AO			IFSA			ARIFPD_AO		
			Acc	$\|B\|$	Acc	Time	$\|B\|$	Acc	Time	$\|B\|$	Acc	Time
3	Ionosphere	$\|U_1\| = 204$	0.774	11	**0.833**	0.001	10	**0.833**	0.498	**7**	**0.833**	0.001
		$\|U_2\| = 233$	0.785	14	0.837	0.040	12	0.799	0.399	**8**	**0.876**	0.102
		$\|U_3\| = 262$	0.805	14	0.832	0.001	12	0.813	0.125	**8**	**0.863**	0.001
		$\|U_4\| = 291$	0.818	14	0.856	0.001	13	0.790	0.242	**8**	**0.866**	0.001
		$\|U_5\| = 320$	0.834	14	0.869	0.001	16	0.834	0.364	**8**	**0.875**	0.001
		$\|U_6\| = 351$	0.843	14	**0.883**	0.001	16	0.832	0.016	**8**	0.880	0.001
4	Pc4	$\|U_1\| = 850$	0.884	7	0.864	0.003	10	0.875	2.995	**2**	**0.884**	0.010
		$\|U_2\| = 971$	0.882	7	0.868	0.003	10	0.864	0.044	**2**	**0.884**	0.019
		$\|U_3\| = 1092$	0.878	7	0.867	0.004	10	0.868	1.086	**2**	**0.881**	0.016
		$\|U_4\| = 1213$	0.888	7	0.876	0.006	11	0.879	2.461	**2**	**0.883**	0.020
		$\|U_5\| = 1334$	0.883	7	0.869	0.006	11	0.873	0.067	**2**	**0.879**	0.025
		$\|U_6\| = 1458$	0.883	7	0.871	0.009	12	**0.880**	0.086	**2**	0.878	0.034
5	Spectf	$\|U_1\| = 203$	0.764	11	**0.793**	0.034	5	0.763	0.119	**2**	**0.793**	0.302
		$\|U_2\| = 232$	0.772	16	0.780	0.002	5	0.780	0.001	**2**	**0.827**	0.001
		$\|U_3\| = 261$	0.774	16	0.797	0.002	6	0.793	0.160	**3**	**0.812**	0.001
		$\|U_4\| = 290$	0.766	16	0.786	0.064	6	0.779	0.005	**3**	**0.797**	0.016
		$\|U_5\| = 319$	0.790	16	0.787	0.004	6	0.771	0.007	**3**	**0.809**	0.001
		$\|U_6\| = 349$	0.742	16	0.742	0.005	6	0.734	0.007	**3**	**0.771**	0.001
6	Movement	$\|U_1\| = 210$	0.810	28	0.790	0.001	9	**0.795**	7.178	**8**	0.781	0.001
		$\|U_2\| = 240$	0.717	28	0.679	0.001	11	**0.708**	4.504	**8**	**0.708**	0.001
		$\|U_3\| = 270$	0.659	28	0.626	0.001	12	0.652	0.017	**8**	**0.678**	0.001
		$\|U_4\| = 300$	0.663	28	0.623	0.001	14	0.637	0.833	**8**	**0.680**	0.133
		$\|U_5\| = 330$	0.639	28	0.615	0.001	14	0.615	0.029	**9**	**0.633**	0.001
		$\|U_6\| = 360$	0.644	28	0.608	0.001	**14**	0.597	0.025	30	**0.622**	6.637
7	Ozone	$\|U_1\| = 1478$	0.918	12	**0.917**	0.006	7	0.915	11.72	**2**	**0.917**	0.020
		$\|U_2\| = 1689$	0.923	19	0.923	11.97	8	0.922	8.934	**2**	**0.925**	0.045
		$\|U_3\| = 1900$	0.927	19	0.928	0.015	8	0.928	20.82	**2**	**0.930**	0.044
		$\|U_4\| = 2111$	0.928	72	0.928	76.06	8	0.927	0.239	**2**	**0.929**	0.057
		$\|U_5\| = 2322$	0.931	72	0.931	0.019	8	0.932	0.233	**2**	**0.936**	0.090
		$\|U_6\| = 2534$	0.933	72	0.933	0.027	8	0.934	0.283	**2**	**0.936**	0.130
8	ORL	$\|U_1\| = 1478$	0.718	158	**0.776**	13.19	77	**0.776**	76.31	**72**	0.764	10.75
		$\|U_2\| = 1689$	0.738	159	0.778	0.694	88	**0.789**	118.8	**73**	0.748	2.364
		$\|U_3\| = 1900$	0.766	160	0.769	0.852	93	0.762	44.55	**73**	**0.773**	0.001
		$\|U_4\| = 2111$	0.759	163	0.768	4.141	**98**	**0.807**	68.08	152	**0.807**	318.6
		$\|U_5\| = 2322$	0.754	173	0.784	19.65	**107**	0.778	174.5	152	**0.800**	0.002
		$\|U_6\| = 2534$	0.770	185	0.817	27.47	**124**	0.800	321.5	152	**0.822**	0.002
9	Mfeat	$\|U_1\| = 1166$	0.901	52	**0.913**	3.989	18	0.702	14.11	**10**	0.905	10.58
		$\|U_2\| = 1332$	0.894	53	0.915	0.550	19	0.699	2.313	**13**	**0.921**	14.68
		$\|U_3\| = 1498$	0.877	54	0.905	0.639	19	0.692	2.650	**14**	**0.909**	4.082
		$\|U_4\| = 1664$	0.880	54	0.904	0.025	18	0.701	3.372	**15**	**0.910**	5.255
		$\|U_5\| = 1830$	0.874	54	0.895	0.020	20	0.738	28.84	**15**	**0.897**	0.062
		$\|U_6\| = 2000$	0.809	54	0.820	0.030	**21**	0.716	0.078	25	**0.830**	23.98
10	Wall-robot	$\|U_1\| = 3182$	0.676	17	0.654	0.045	20	0.666	9.436	**4**	**0.738**	0.139
		$\|U_2\| = 3636$	0.697	18	0.671	1.204	22	0.668	1.574	**5**	**0.723**	8.218
		$\|U_3\| = 4090$	0.730	18	0.711	0.074	23	0.721	0.947	**5**	**0.765**	0.283
		$\|U_4\| = 4544$	0.753	18	0.732	0.167	23	0.749	1.031	**5**	**0.770**	0.439
		$\|U_5\| = 4998$	0.757	18	0.749	0.174	23	0.754	1.147	**5**	**0.796**	0.498
		$\|U_6\| = 5456$	0.759	18	0.758	0.248	23	0.758	1.579	**5**	**0.811**	0.608

(Continued)

Table 3 (continued)

ID	Data sets	Adding data sets	RAW	FDAR_AO			IFSA			ARIFPD_AO		
			Acc	$\|B\|$	Acc	Time	$\|B\|$	Acc	Time	$\|B\|$	Acc	Time
11	Hill-valley	$\|U_1\| = 707$	0.518	4	0.505	0.560	13	0.492	3.157	**2**	**0.513**	0.006
		$\|U_2\| = 808$	0.522	4	0.488	0.003	15	0.486	2.627	**2**	**0.505**	0.011
		$\|U_3\| = 909$	0.514	4	0.503	0.004	16	0.485	6.192	**2**	**0.514**	0.012
		$\|U_4\| = 1010$	0.510	4	0.490	0.008	19	0.472	5.273	**2**	**0.510**	0.019
		$\|U_5\| = 1111$	0.499	4	0.463	0.010	20	0.469	0.053	**2**	**0.499**	0.033
		$\|U_6\| = 1212$	0.515	4	0.465	0.013	20	0.488	0.050	**2**	**0.515**	0.040

We consider 66 cases when adding the set of objects. There are 7 cases in which our reducts have no higher classification accuracy than the reducts of FDAR_AO. In the remaining 59 cases, the reducts of the proposed method show superiority in accuracy compared to FDBAR_AO and IFSA. Clearly, the proposed incremental method is also very effective on datasets with low initial classification accuracy. In other words, the attribute reduction methods based on the rough set approach and its extensions have many difficulties in improving the classification accuracy for the noisy data.

Based on the accuracy of algorithms in Table 3, paired two-tailed t-tests were also conducted with a confidence level of 0.95 to evaluate the differences between FDAR_AO, IFSA, and ARIFPD_AO. The corresponding *p*-values (two-tailed) were found at levels 1.17E-10 and 6.85E-10 for KNN. These results provide a sound basis to conclude that our algorithm outperforms the compared algorithms in statistical significance. From the above results, it can be confirmed that the incremental attribute reduction algorithm using the intuitionistic fuzzy set approach has outstanding advantages compared to algorithms based on the fuzzy rough set when processing noisy and inconsistent data.

5 Conclusion

With the primary purpose of reducing the number of features and improving the classification ability, attribute reduction is considered as a critical problem in the data preprocessing step. This paper recommends a measure for intuitionistic fuzzy partition distance and constructs an incremental formula to update the intuitionistic fuzzy partition distance when adding an object set. Thereby, this paper constructs two algorithms based on the intuitionistic fuzzy set approach. The first algorithm proposed to find the reduct on the decision table when there is no additional set of objects. The second algorithm is the incremental algorithm to find the approximate reduct when the decision table is increased in the object set. Compared to methods based on the rough set and fuzzy rough set approaches, the experimental results have shown that our methods can ameliorate the accuracy of inconsistent or low initial classification accuracy data sets.

It can be easily seen that the limitation of intuitionistic fuzzy set-based algorithms is the execution time due to the supplement of the diversity degree in intuitionistic fuzzy equivalence classes. For future researchers, we focus on developing the incremental formula to reduce the computational time and guarantee the classification accuracy. In addition, we also design incremental algorithms based on the intuitionistic fuzzy sets using granular structures. We will continue to study the incremental methods that find reducts from the decision tables in the case of supplementing and removing the set of attributes.

Acknowledgement: The authors thank you for the generous support of the Simulation and High-Performance Computing department (SHPC) of the HaUI Institute of Technology (HIT) for conducting experiments in this paper and sincerely thank you for the assistance of the research team VAST.

Funding Statement: This research is funded by Hanoi University of Industry under Grant Number 27-2022-RD/HĐ-ĐHCN (URL: https://www.haui.edu.vn/).

Author Contributions: Study conception and design: Pham Viet Anh, Nguyen The Thuy; data collection: Pham Dinh Khanh; analysis and interpretation of results: Nguyen Long Giang, Nguyen Ngoc Thuy; draft manuscript preparation: Pham Viet Anh, Nguyen The Thuy, Nguyen Long Giang, Nguyen Ngoc Thuy. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data presented in this study can be made available upon reasonable request from the corresponding author.

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

References

[1] X. Zhang, C. L. Mei, D. G. Chen and Y. Y. Yang, "A fuzzy rough set-based feature selection method using representative instances," *Knowledge Based Systems*, vol. 151, pp. 216–229, 2018. https://doi.org/10.1016/j.knosys.2018.03.031

[2] T. K. Sheeja and A. S. Kuriakose, "A novel feature selection method using fuzzy rough sets," *Computers in Industry*, vol. 97, no. 3, pp. 111–116, 2018. https://doi.org/10.1016/j.compind.2018.01.014

[3] X. Che, D. Chen and J. Mi, "Label correlation in multi-label classification using local attribute reductions with fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 426, pp. 121–144, 2022. https://doi.org/10.1016/j.fss.2021.03.016

[4] Z. Huang and J. Li, "A fitting model for attribute reduction with fuzzy β-covering," *Fuzzy Sets and Systems*, vol. 413, pp. 114–137, 2021. https://doi.org/10.1016/j.fss.2020.07.010

[5] Q. H. Hu, D. R. Yu and Z. X. Xie, "Information-preserving hybrid data reduction based on fuzzy-rough techniques," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 414–423, 2016.

[6] A. Meriello and R. Battiti, "Feature selection based on the neighborhood entropy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 12, pp. 6313–6322, 2018.

[7] B. Sang, H. Chen, L. Yang, T. Li and W. Xu, "Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 6, pp. 1683–1697, 2021.

[8] Z. Yuan, H. Chen and T. Li, "Exploring interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data," *Pattern Recognition*, vol. 127, pp. 108651, 2022. https://doi.org/10.1016/j.patcog.2022.108651

[9] C. Z. Wang, Y. Huang, M. W. Shao and X. D. Fan, "Fuzzy rough set based attribute reduction using distance measures," *Knowledge-Based Systems*, vol. 164, pp. 205–212, 2019. https://doi.org/10.1016/j.knosys.2018.10.038

[10] N. L. Giang, L. H. Son, T. T. Ngan, T. M. Tuan, H. T. Phuong *et al.,* "Novel incremental algorithms for attribute reduction from dynamic decision systems using hybrid filter-wrapper with fuzzy partition distance," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 858–873, 2020.

[11] Y. M. Liu, S. Y. Zhao, H. Chen, C. P. Li and Y. M. Lu, "Fuzzy rough incremental attribute reduction applying dependency measures,," *APWeb-WAIM 2017 Web and Big Data*, vol. 10366, pp. 484–492, 2017.

[12] Y. Y. Yang, D. G. Chen, H. Wang, E. C. 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. https://doi.org/10.1016/j.fss.2016.08.001

[13] Y. Y. Yang, D. G. Chen, H. Wang and X. H. Wang, "Incremental perspective for feature selection based on fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1257–1273, 2017.

[14] 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.

[15] P. Ni, S. Y. Zhao, X. H. Wang, H. Chen, C. P. Li *et al.,* "Incremental feature selection based on fuzzy rough sets," *Information Sciences*, vol. 536, no. 5, pp. 185–204, 2020. https://doi.org/10.1016/j.ins.2020.04.038

[16] N. L. Giang, L. H. Son, N. A. Tuan, T. T. Ngan, N. N. Son *et al.,* "Filter-wrapper incremental algorithms for finding reduct in incomplete decision systems when adding and deleting an attribute set," *International Journal of Data Warehousing and Mining*, vol. 17, no. 2, pp. 39–62, 2021.

[17] N. T. Thang, N. L. Giang, H. V. Long, N. A. Tuan, T. M. Tuan *et al.,* "Efficient algorithms for dynamic incomplete decision systems," *International Journal of Data Warehousing and Mining*, vol. 17, no. 3, pp. 44–67, 2021.

[18] W. L. Hung and M. S. Yang, "Similarity measures of intuitionistic fuzzy sets based on Lp metric," *International Journal of Approximate Reasoning*, vol. 46, no. 1, pp. 120–136, 2007.

[19] B. Huang, C. X. Guo, Y. L. Zhuang, H. X. Li and X. Z. Zhou, "Intuitionistic fuzzy multigranulation rough sets," *Information Sciences*, vol. 277, pp. 299–320, 2014. https://doi.org/10.1016/j.ins.2014.02.064

[20] A. Tan, S. Shi, W. Z. Wu, J. Li, W. Pedrycz *et al.,* "Granularity and entropy of intuitionistic fuzzy information and their applications," *IEEE Transactions on Cybernetics*, vol. 52, no. 1, pp. 192–204, 2022.

[21] N. T. Thang, N. L. Giang, T. T. Dai, T. T. Tuan, N. Q. Huy *et al.,* "A novel filter-wrapper algorithm on intuitionistic fuzzy set for attribute reduction from decision tables," *International Journal of Data Warehousing and Mining*, vol. 17, no. 4, pp. 67–100, 2021. https://doi.org/10.4018/IJDWM

[22] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[23] I. Iancu, "Intuitionistic fuzzy similarity measures based on Frank t-norms family," *Pattern Recognition Letters*, vol. 42, pp. 128–136, 2014. https://doi.org/10.1016/j.patrec.2014.02.010

[24] W. Shu, W. Qian and Y. Xie, "Incremental approaches for feature selection from dynamic data with the variation of multiple objects," *Knowledge-Based Systems*, vol. 163, pp. 320–331, 2019. https://doi.org/10.1016/j.knosys.2018.08.028

[25] UCI Machine Learning Repository, "Data," 2021. [Online]. Available: https://archive.ics.uci.edu/ml/index.php