



ARTICLE

## Attribute Reduction of Hybrid Decision Information Systems Based on Fuzzy Conditional Information Entropy

Xiaoqin Ma<sup>1,2</sup>, Jun Wang<sup>1</sup>, Wenchang Yu<sup>1</sup> and Qinli Zhang<sup>1,2,\*</sup>

<sup>1</sup>School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, 247000, China

<sup>2</sup>Anhui Education Big Data Intelligent Perception and Application Engineering Research Center, Chizhou, 247000, China

\*Corresponding Author: Qinli Zhang. Email: zqlynu@163.com

Received: 28 December 2023 Accepted: 13 March 2024 Published: 15 May 2024

### ABSTRACT

The presence of numerous uncertainties in hybrid decision information systems (HDISs) renders attribute reduction a formidable task. Currently available attribute reduction algorithms, including those based on Pawlak attribute importance, Skowron discernibility matrix, and information entropy, struggle to effectively manage multiple uncertainties simultaneously in HDISs like the precise measurement of disparities between nominal attribute values, and attributes with fuzzy boundaries and abnormal values. In order to address the aforementioned issues, this paper delves into the study of attribute reduction within HDISs. First of all, a novel metric based on the decision attribute is introduced to solve the problem of accurately measuring the differences between nominal attribute values. The newly introduced distance metric has been christened the supervised distance that can effectively quantify the differences between the nominal attribute values. Then, based on the newly developed metric, a novel fuzzy relationship is defined from the perspective of “feedback on parity of attribute values to attribute sets”. This new fuzzy relationship serves as a valuable tool in addressing the challenges posed by abnormal attribute values. Furthermore, leveraging the newly introduced fuzzy relationship, the fuzzy conditional information entropy is defined as a solution to the challenges posed by fuzzy attributes. It effectively quantifies the uncertainty associated with fuzzy attribute values, thereby providing a robust framework for handling fuzzy information in hybrid information systems. Finally, an algorithm for attribute reduction utilizing the fuzzy conditional information entropy is presented. The experimental results on 12 datasets show that the average reduction rate of our algorithm reaches 84.04%, and the classification accuracy is improved by 3.91% compared to the original dataset, and by an average of 11.25% compared to the other 9 state-of-the-art reduction algorithms. The comprehensive analysis of these research results clearly indicates that our algorithm is highly effective in managing the intricate uncertainties inherent in hybrid data.

### KEYWORDS

Hybrid decision information systems; fuzzy conditional information entropy; attribute reduction; fuzzy relationship; rough set theory (RST)



## Symbols and Notations Used in the Paper

Symbols and notations	Meaning
$X$	A set of objects
$C$	A set of attributes
$BA$	Set of conditional attributes
$D$	The decision attribute
$2^X$	The power set of $X$
$ Y $	The cardinality of $Y$
$I$	$[0,1]$
$I^X$	A set of all fuzzy sets on $X$
$I^{X \times X}$	A set of all fuzzy relationship on $X$
$R$	A fuzzy relationship
$R^x$	The fuzzy information granule of $x$
$\theta$	The threshold value of tolerance relationship
$P$	A subset of conditional attributes
$T_p^\theta$	A tolerance relationship determined by $\theta$ and $P$

## 1 Introduction

### 1.1 Research Background

In information systems, the attributes of data are not of equal importance. Some of them may not have a direct or significant impact on decision-making. These attributes are considered as redundant. The objective of attribute reduction is to eliminate these redundant attributes, while maintaining the same classification accuracy, to facilitate more concise and efficient decision-making. Attribute reduction is also called feature selection.

With the emergence of the era of big data, the intricacies and diversities of data have been constantly on the rise. Many datasets encompass both nominal and real-valued attributes, which are referred to as hybrid data. Such hybrid data can be found in various domains, including finance, healthcare, and e-commerce. The hybrid data encompass diverse attributes, necessitating distinct methodologies and strategies for their manipulation. The complexity intensifies the challenges of attribute reduction. Primarily, the disparities among nominal attribute values prove challenging to assess accurately using Euclidean distance. Furthermore, data may encompass uncertainties such as fuzzy attributes and anomalous attribute values. The reduction of attributes in hybrid data is a pivotal research topic in the field of data mining and lies at the heart of rough set theory. This line of research not only enhances the efficiency and precision of data processing but also fosters the advancement of data mining and machine learning. Consequently, exploring the attribute reduction problem in mixed information systems holds profound theoretical significance and practical relevance [1].

### 1.2 Related Works

The classical rough set theory is limited in its ability to handle continuous data, hence it has been enhanced to include neighborhood rough sets and fuzzy rough sets as extensions. Currently, numerous scholars have delved into the intricacies of attribute reduction by using neighborhood rough sets and fuzzy rough sets, ultimately yielding impressive outcomes.

Neighborhood rough sets can effectively handle continuous values and have been widely applied in attribute reduction. Fan et al. [2] established a max-decision neighborhood rough set model and

successfully applied it to achieve attribute reduction for information systems, leading to more accurate and concise decision-making. Shu et al. [3] proposed a neighborhood entropy-based incremental feature selection framework for hybrid data based on the neighborhood rough set model. In their study, Zhang et al. [4] conducted the entropy measurement on the gene space, utilizing the neighborhood rough sets, for effective gene selection. Zhou et al. [5] put forward an online stream-based attribute reduction method based on the new neighborhood rough set theory, eliminating the need for domain knowledge or pre-parameters. Xia et al. [6] built a granular ball neighborhood rough set method for fast adaptive attribute reduction. Liao et al. [7] developed an effective attribute reduction method for handling hybrid data under test cost constraints based on the generalized confidence level and vector neighborhood rough set models. None of the above methods provides a specific strategy for handling the differences between nominal attribute values. Instead, they use Euclidean distance to measure the differences between nominal attribute values, or use category information to simply define the differences between nominal attribute values as 1 or 0.

Fuzzy rough sets are capable of effectively dealing with the ambiguity in real-world scenarios, and have also been successfully applied in attribute reduction. Yuan et al. [8] established a unsupervised attribute reduction model for hybrid data based on fuzzy rough sets. Zeng et al. [9] examined the incremental updating mechanism of fuzzy rough approximations in response to attribute value changes in a hybrid information system. Yang et al. [10] considered the uncertainty measurement and attribute reduction for multi-source fuzzy information systems based on a new multi-granulation rough sets model. Wang et al. [11] utilized variable distance parameters and, drawing from fuzzy rough set theory, developed an iterative attribute reduction model. Singh et al. [12] came up with a fuzzy similarity-based rough set approach for attribute reduction in set-valued information systems; Jain et al. [13] created an attribute reduction model by using the intuitionistic fuzzy rough set. The above methods are either not suitable for handling mixed data or do not have special techniques for handling nominal attribute values and abnormal attribute values.

Information entropy quantifies the uncertainty of knowledge and can be used for the attribute reduction. Li et al. [14] used conditional information entropy to measure the uncertainty and reduce the attributes of multi-source incomplete information systems. Vergara et al. [15] summarized the feature selection methods based on the mutual information. Li et al. [16] measured the uncertainty of gene spaces and designed a gene selection algorithm using information entropy. Zhang et al. [17] defined a fuzzy information entropy of classification data based on the constructed fuzzy information structure and used it for attribute reduction. Zhang et al. [18] constructed a hybrid data attribute reduction model utilizing the  $\lambda$ -conditional entropy derived from fuzzy rough set theory. Sang et al. [19] proposed an incremental attribute reduction method using the fuzzy dominance neighborhood conditional entropy derived from fuzzy dominant neighborhood rough sets. Huang et al. [20] defined a new fuzzy conditional information entropy for fuzzy  $\beta$  covering information systems and applied it to attribute reduction. The above methods also do not provide a specific solution for handling mixed data.

Akram et al. [21,22] presented the concept of attribute reduction and designed the associated attribute reduction algorithms based on the discernibility matrix and discernibility function. These methods are intuitive and easy to comprehend, enabling the calculation of all reducts. However, they exhibit a high computational complexity, rendering them unsuitable for large datasets.

The comprehensive overview of the strengths and weaknesses of diverse attribute reduction algorithms can be found in [Table 1](#).

**Table 1:** Details of various attribute reduction algorithms

Algorithms	References	Strengths	Weaknesses
Pawlak attribute importance	[2–6]	Strong interpretability, high stability, no prior knowledge required, and the ability to handle uncertainty.	Susceptibility to outliers and noise, challenges in handling large-scale datasets, and the inability to effectively leverage prior knowledge.
Information entropy	[14–17]	Strong interpretability, wide applicability, and effective avoidance of redundant attributes.	Extensive computational requirements, susceptibility to outliers, and inability to guarantee global optimality.
Skowron discernibility matrix	[21,22]	Concise and straightforward expression, ease of operation, and the ability to derive all possible reductions.	large storage space requirements, low computational efficiency, sensitivity to outliers and noise, and unsuitability for continuous attributes.

### 1.3 Motivation and Contribution

Hybrid data is a common occurrence in data mining. They have been observed that nominal attributes significantly impact the measurement of data similarity, and Euclidean distance do not work well when dealing with nominal attribute values, yet the majority of current reduction algorithms rely heavily on Euclidean distance. Therefore, the supervised distance, a novel metric specifically for measuring the difference between nominal attribute values, has been meticulously crafted. Existing attribute reduction algorithms often exhibit sensitivity to abnormal attribute values and lack an efficient mechanism to address them. Consequently, this article introduces a novel fuzzy relationship that replaces similarity measurements based on attribute values with the count of similar attributes. This innovative approach effectively filters out abnormal attribute values, enhancing the overall accuracy and robustness of the algorithm. Due to the fact that the fuzzy conditional information entropy not only considers the fuzziness of data but also has a certain anti-noise ability, this paper uses fuzzy conditional information entropy to reduce the attributes of mixed data. The article's novelty and contribution are outlined below:

1. A new distance metric called the supervised distance is introduced. The supervised distance metric, which considers the decision attribute that affect attribute similarity, leads to more precise attribute reduction.

2. Based on the new distance measurement, a new fuzzy relationship determined by the quantity of attributes with similar values is established. This approach views the relationship from the perspective of “feedback on parity of attribute values to attribute sets”, resulting in a fuzzy relation that is robust to a few abnormal attributes.

3. The fuzzy conditional information entropy is defined based on the new fuzzy relationship and an advanced attribute reduction algorithm is developed based on the fuzzy conditional information

entropy. This algorithm incorporates an innovative distance function, an innovative fuzzy relationship, and fuzzy conditional information entropy.

4. Through meticulous experimental validation, this study clearly demonstrates that the fuzzy relation offers superior performance to the equivalence relation when dealing with hybrid data, effectively handling its complexity and uncertainties.

### 1.4 Organization

The structure of this paper is outlined as follows. Section 2 provides a brief overview of fuzzy relationships and HDISs, laying the foundation for the subsequent sections. Section 3 describes the method in detail. Section 4 presents the experimental results and discusses them. Section 5 summarizes the paper.

## 2 Preliminaries

In this section, we will review some fundamental concepts of fuzzy relations and HDISs.

### 2.1 Fuzzy Relation

Let  $I = [0, 1]$ . A fuzzy set  $F$  on  $X$  is known as a mapping  $\mu: X \rightarrow I$ .  $\forall x \in X$ ,  $\mu(x)$  is the degree of membership of  $x$  in  $F$  [23].  $F$  can be represented as follows:

$$F = \frac{\mu(x_1)}{x_1} + \frac{\mu(x_2)}{x_2} + \dots + \frac{\mu(x_n)}{x_n}. \tag{1}$$

Let  $|F| = \sum_{i=1}^n \mu(x_i)$ . Then  $|F|$  is called the cardinality of  $F$ .

Let  $I^X$  denote the family of all fuzzy sets on  $X$ .

$\forall F \in I^X$  and  $\forall G \in 2^X$ ,  $F \cap G$  is defined as follows:

$$(F \cap G)(x) = F(x) \wedge G(x) = \begin{cases} F(x), & x \in G \\ 0, & x \notin G \end{cases} \tag{2}$$

Then  $|F \cap G| = |F|$ .

If  $R$  is a fuzzy set on  $X \times X$ , then  $R$  is referred to as a fuzzy relation on  $X$ .  $R$  can be represented by the matrix:  $M(R) = (r_{ij})_{n \times n}$  ( $r_{ij} = R(x_i, x_j) \in I$  denotes the similarity between  $x_i$  and  $x_j$ ).

Let  $I^{X \times X}$  denote the set of all fuzzy relationship on  $X$ .

**Definition 2.1.** Reference [24] Let  $R$  be a fuzzy relation on  $X$ . Then  $R$  is

- 1) Reflexive, if  $R(x, x) = 1$  ( $\forall x \in X$ ).
- 2) Symmetric, if  $R(x, y) = R(y, x)$  ( $\forall x, y \in X$ ).
- 3) Transitive, if  $R(x, z) \geq R(x, y) \wedge R(y, z)$  ( $\forall x, y, z \in X$ ).

If  $R$  is reflexive, symmetric, and transitive, then  $R$  is called a fuzzy equivalence relation on  $X$ . If  $R$  is reflexive and symmetric, then  $R$  is called a fuzzy tolerance relation on  $X$ .

Let  $R \in I^{X \times X}$ .  $\forall x \in X$ , the fuzzy set  $R^x$  on  $X$  is defined as follows:

$$R^x(y) = R(x, y) (y \in X). \tag{3}$$

Then  $R^x$  is called the fuzzy information granule of  $x$ .

According to (1),  $R^x = \frac{R(x, x_1)}{x_1} + \frac{R(x, x_2)}{x_2} + \dots + \frac{R(x, x_n)}{x_n}$ .

Then  $|R^x| = \sum_{i=1}^n R(x, x_i)$ .

## 2.2 Hybrid Decision Information Systems

**Definition 2.2.** Reference [25]  $(X, C)$  is called an information system (IS), if  $\forall c \in C$  decides a function  $f: X \rightarrow V_c$ , where  $V_c = \{f(x) : x \in X\}$ . If  $Q \subseteq C$ , then  $(X, Q)$  is said to be a subsystem of  $(X, C)$ .

If  $C = B \cup \{d\}$ , where  $B = \{b_1, b_2, \dots, b_m\}$  represents the conditional attributes set,  $d$  represents the decision attribute, then  $(X, C)$  is called as a decision information system.

**Definition 2.3.** Let  $(X, B, d)$  is a decision information system. Then  $(X, B, d)$  is known as a hybrid decision information system (HDIS), if  $B = B^c \cup B^r$ , where  $B^c$  is a set of categorical attributes,  $B^r$  is a set of real-valued attributes.

**Example 2.4.** Table 2 shows an HDIS, where  $X = \{x_1, x_2, \dots, x_9\}$ ,  $B^c = \{b_1, b_2\}$  and  $B^r = \{b_3\}$ .

**Table 2:** An HDIS

$X$	Headache ( $b_1$ )	Muscle pain ( $b_2$ )	Temperature ( $b_3$ )	Symptom ( $d$ )
$x_1$	*	Yes	37.2	Health
$x_2$	Middle	No	*	Rhinitis
$x_3$	Middle	*	39.1	Flu
$x_4$	No	*	*	Health
$x_5$	*	Yes	36.5	Rhinitis
$x_6$	Sick	Yes	39.8	Flu
$x_7$	No	No	36.0	Health
$x_8$	No	Yes	38.3	Flu
$x_9$	Sick	Yes	39.8	Flu

Note: \* denotes missing values.

## 3 Methodology

### 3.1 A New Distance Function

To accurately measure the difference between two attributes in HDISs, we have developed a novel distance function that accounts for various types of attributes and missing data. This innovative approach allows for a more comprehensive assessment of the similarity or dissimilarity between different attributes in the system.

We first assign a definition for the distance between categorical attribute values to enable further distance definition of hybrid data.

**Definition 3.1.** For an HDIS  $(X, B, d)$ , let  $V_d = \{d(x) : x \in X\} = \{d_1, d_2, \dots, d_r\}$ ,

$$N(b, x) = |\{y \in X : b(x) = b(y), b \in B^c, b(x) \neq *\}| (\forall x \in X)$$

$$N_i(b, x) = |\{y \in X : b(x) = b(y), d(y) = d_i\}|$$

$$\text{Then } N(b, x) = \sum_{i=1}^r N_i(b, x).$$

**Example 3.2.** (Continue with Example 2.4)

$$\begin{aligned} V_d &= \{d_1 = Flu, d_2 = Rhinitis, d_3 = Health\}, N(b_2, x_5) = |\{x \in X : b_2(x) = b_2(x_5) = Yes\}| \\ &= |\{x_1, x_2, x_4, x_5, x_9\}| = 5, N_1(b_2, x_5) = |\{x \in X : b_2(x) = b_2(x_5) = Yes, d(x) = d_1 = Flu\}| \\ &= |\{x_1, x_2, x_4\}| = 3, N_2(b_2, x_5) = |\{x \in X : b_2(x) = b_2(x_5) = Yes, d(x) = d_2 = Rhinitis\}| = |\{x_5\}| \\ &= 1, N_3(b_2, x_5) = |\{x \in X : b_2(x) = b_2(x_5) = Yes, d(x) = d_3 = Health\}| = |\{x_9\}| = 1. \end{aligned}$$

**Definition 3.3.** For an HDIS  $(X, B, d)$ , let  $|V_d| = r, b \in B^c, x \in X, y \in X, b(x) \neq * \text{ and } b(y) \neq *$ . Then the distance between  $b(x)$  and  $b(y)$  is defined as follows:

$$\rho_c(b(x), b(y)) = \frac{1}{2} \sum_{i=1}^r \left| \frac{N_i(b, x)}{N(b, x)} - \frac{N_i(b, y)}{N(b, y)} \right|. \tag{4}$$

The distribution of attribute values defines a distance that is well-suited for the characteristics of classification data.

**Proposition 3.4.** Let  $(X, B, d)$  be an HDIS. Then the following conclusions are valid:

- 1)  $\rho_c(b(x), b(x)) = 0$ ,
- 2)  $0 \leq \rho_c(b(x), b(y)) \leq 1$ .

*Proof.* 1) The conclusion is self-evidently valid.

2)  $\rho_c(b(x), b(y)) \geq 0$  is self-evidently valid.

Since

$$\sum_{i=1}^r \frac{N_i(b, x)}{N(b, x)} = \sum_{i=1}^r \frac{N_i(b, y)}{N(b, y)} = 1,$$

we have

$$\rho_c(b(x), b(y)) \leq \frac{1}{2} \left( \sum_{i=1}^r \frac{N_i(b, x)}{N(b, x)} + \sum_{i=1}^r \frac{N_i(b, y)}{N(b, y)} \right) = 1.$$

**Definition 3.5.** For an HDIS  $(X, B, d)$ , let  $b \in B^r, x \in X, y \in X, b(x) \neq *, \text{ and } b(y) \neq *$ . Then the distance between  $b(x)$  and  $b(y)$  is defined as follows:

$$\rho_r(b(x), b(y)) = \frac{|b(x) - b(y)|}{\max\{b(x) : x \in X\} - \min\{b(x) : x \in X\}}. \tag{5}$$

$\rho_r(b(x), b(x)) = 0$  and  $0 \leq \rho_r(b(x), b(y)) \leq 1$  are self-evidently valid.

**Definition 3.6.** For an HDIS  $(X, B, d)$ , let  $b \in B, x \in X, \text{ and } y \in X$ . Then the distance between  $b(x)$  and  $b(y)$  is defined as follows:

$$\rho(b(x), b(y)) = \begin{cases} 0, & b \in B, b(x) = * \vee b(y) = *, d(x) = d(y) \\ 1, & b \in B, b(x) = * \vee b(y) = *, d(x) \neq d(y) \\ \rho_c(b(x), b(y)), & b \in B^c, b(x) \neq * \wedge b(y) \neq * \\ \rho_r(b(x), b(y)), & b \in B^r, b(x) \neq * \wedge b(y) \neq * \end{cases}. \tag{6}$$

**Example 3.7.** (Continuation of Example 2.4) According to Definitions 3.1–3.6, we have

$$1) \rho(b_1(x_1), b_1(x_3)) = \left( \left| \frac{2}{2} - \frac{1}{2} \right| + \left| \frac{0}{2} - \frac{1}{2} \right| + \left| \frac{0}{2} - \frac{0}{2} \right| \right) / 3 = \frac{1}{3},$$

$$2) \rho(b_1(x_1), b_1(x_4)) = \left( \left| \frac{2}{2} - \frac{1}{3} \right| + \left| \frac{0}{2} - \frac{0}{3} \right| + \left| \frac{0}{2} - \frac{2}{3} \right| \right) / 3 = \frac{4}{9},$$

$$3) \rho(b_1(x_5), b_1(x_9)) = 1,$$

$$4) \rho(b_2(x_1), b_2(x_3)) = 0,$$

$$5) \rho(b_3(x_1), b_3(x_3)) = \frac{|39.5 - 39|}{40 - 36} = \frac{1}{8}.$$

**Definition 3.8.** For an HDIS  $(X, B, d)$ ,  $\forall b \in B, \forall x_i, x_j \in X$ , let  $M_b = (\rho(b(x_i), b(x_j)))_{n \times n}$ . Then  $M_b$  is called the distance matrix of  $b$ .

**Example 3.9.** (Continuation of Example 2.4) Here are the distance matrices of  $b_1$  and  $b_3$ :

$$M_{b_1} = \begin{pmatrix} 0 & 0 & 1/3 & 4/9 & 1 & 1/3 & 4/9 & 4/9 & 1 \\ 0 & 0 & 1/3 & 4/9 & 1 & 1/3 & 4/9 & 4/9 & 1 \\ 1/3 & 1/3 & 0 & 4/9 & 1 & 0 & 4/9 & 4/9 & 1 \\ 4/9 & 4/9 & 4/9 & 0 & 1 & 4/9 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1/3 & 1/3 & 0 & 4/9 & 0 & 0 & 4/9 & 4/9 & 1 \\ 4/9 & 4/9 & 4/9 & 0 & 1 & 4/9 & 0 & 0 & 0 \\ 4/9 & 4/9 & 4/9 & 0 & 1 & 4/9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix},$$

$$M_{b_3} = \begin{pmatrix} 0 & 1/8 & 1/8 & 3/8 & 3/4 & 1 & 7/8 & 1 & 5/8 \\ 1/8 & 0 & 1/4 & 1/2 & 7/8 & 1 & 1 & 1 & 3/4 \\ 1/8 & 1/4 & 0 & 1/4 & 5/8 & 1 & 3/4 & 1 & 1/2 \\ 3/8 & 1/2 & 1/4 & 0 & 3/8 & 1 & 1/2 & 1 & 1/4 \\ 3/4 & 7/8 & 5/8 & 3/8 & 0 & 0 & 1/8 & 1 & 1/8 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 7/8 & 1 & 3/4 & 1/2 & 1/8 & 1 & 0 & 0 & 1/4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 5/8 & 1 & 3/4 & 1/2 & 1/4 & 1 & 1/4 & 0 & 0 \end{pmatrix}.$$

### 3.2 A New Fuzzy Relation

**Definition 3.10.** For an HDIS  $(X, B, d)$ ,  $\forall P \subseteq B, \forall \theta \in [0, 1]$ , the tolerance relationship is defined as follows:

$$T_p^\theta = \{(x, y) \in X \times X : \forall b \in P, \rho(b(x), b(y)) \leq \theta\}.$$

Let  $T_p^\theta(x) = \{y \in X : (x, y) \in T_p^\theta\}$ .  $T_p^\theta(x)$  is called the tolerance class of  $x$ .

$T_p^\theta$  uses distance to measure the similarity of objects. Next, we introduce a new fuzzy relationship that evaluates the similarity of objects based on the number of similar attributes.

**Definition 3.11.** Let  $(X, B, d)$  be an HDIS.  $\forall P \subseteq B, \forall \theta \in [0, 1], \forall x \in X, \forall y \in X$ , define

$$R_p^\theta(x, y) = \frac{1}{|B|} |\{b \in P : \rho(b(x), b(y)) \leq \theta\}|, R_d = \{(x, y) \in X \times X : d(x) = d(y)\}$$

Then  $R_p^\theta$  and  $R_d$  are a fuzzy relation and an equivalence relation on  $X$ , respectively.

The matrix representation of  $R_p^\theta$  is  $M(R_p^\theta) = (R_p^\theta(x_i, x_j))_{n \times n}$ .

Let  $R_p^{\theta, x}(y) = R_p^\theta(x, y)$  and  $R_d^x = \{y \in X : (x, y) \in R_d\}$ .

$R_p^{\theta, x}$  is the fuzzy information granule of  $x$  and  $R_d^x$  is the decision class of  $x$ .

According to (1),  $R_p^{\theta, x} = \frac{R_p^\theta(x, x_1)}{x_1} + \frac{R_p^\theta(x, x_2)}{x_2} + \dots + \frac{R_p^\theta(x, x_n)}{x_n}$ ,  $|R_p^{\theta, x}| = \sum_{i=1}^n R_p^\theta(x, x_i)$

Let  $X/R_d = \{R_d^x : x \in X\} = \{D_1, D_2, \dots, D_r\}$ .

Let  $P^{\theta, xy} = \{b \in P : \rho(b(x), b(y)) \leq \theta\}$ . Then  $R_p^{\theta, x}(y) = \frac{1}{|B|} |P^{\theta, xy}|$ .

$R_p^{\theta, x}(x) = \frac{|P|}{|B|} \leq 1$  and  $R_p^{\theta, x}(y) \leq \frac{|P|}{|B|}$  are self-evidently valid.

**Proposition 3.12.** *Let  $(X, B, d)$  be an HDIS. If  $P \subseteq B$ ,  $P_1 \subseteq P_2 \subseteq B$ , and  $0 \leq \theta_1 \leq \theta_2 \leq 1$ ,  $R_{P_1}^{\theta_1, x} \subseteq R_{P_2}^{\theta_2, x}$  and  $R_{P_1}^{\theta_1, x} \subseteq R_{P_2}^{\theta_2, x} (\forall x \in X)$ .*

*Proof.* According to Definition 3.11,  $\forall y \in X, \forall \theta \in [0, 1], R_{P_1}^{\theta_1, x}(y) = \frac{1}{|B|} |P_1^{\theta_1, xy}|, R_{P_2}^{\theta_2, x}(y) = \frac{1}{|B|} |P_2^{\theta_2, xy}|$ .

Since  $P_1 \subseteq P_2, P_1^{\theta_1, xy} \subseteq P_2^{\theta_2, xy}$ . Therefore,  $\forall y \in X, R_{P_1}^{\theta_1, x}(y) \leq R_{P_2}^{\theta_2, x}(y)$ .

Thus,  $\forall x \in X, R_{P_1}^{\theta_1, x} \subseteq R_{P_2}^{\theta_2, x}$ .

Since  $0 \leq \theta_1 \leq \theta_2 \leq 1, P_1^{\theta_1, xy} \subseteq P_2^{\theta_2, xy}$ . Thus  $R_{P_1}^{\theta_1, x}(y) \leq R_{P_2}^{\theta_2, x}(y)$ . Therefore,  $R_{P_1}^{\theta_1, x} \subseteq R_{P_2}^{\theta_2, x}$ .

### 3.3 Fuzzy Conditional Information Entropy in HDISs

This section explores the concept of fuzzy entropy measures in HDISs to measure the uncertainty of HDISs.

**Definition 3.13.**  *$(X, B, d)$  is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . Then the fuzzy information entropy of  $P$  is defined as follows:*

$$H^\theta(P) = - \sum_{i=1}^n \frac{|R_P^{\theta, x_i}|}{n} \log_2 \frac{|R_P^{\theta, x_i}|}{n}. \tag{7}$$

**Proposition 3.14.**  *$(X, B, d)$  is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . The following inequality holds:*

$$0 \leq H^\theta(P) \leq n \log_2 \frac{mn}{|P|}.$$

*Proof.* Since  $\frac{|P|}{m} \leq |R_P^{\theta, x_i}| \leq n, \frac{|P|}{mn} \leq \frac{|R_P^{\theta, x_i}|}{n} \leq 1$ .

Thus,  $0 \leq -\log_2 \frac{|R_P^{\theta, x_i}|}{n} \leq \log_2 \frac{mn}{|P|}$ . Hence,  $0 \leq -\frac{|R_P^{\theta, x_i}|}{n} \log_2 \frac{|R_P^{\theta, x_i}|}{n} \leq \log_2 \frac{mn}{|P|}$ .

The following conclusion follows directly from Definition 3.13:

$$0 \leq H^\theta(P) \leq n \log_2 \frac{mn}{|P|}.$$

**Definition 3.15.**  $(X, B, d)$  is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . The fuzzy conditional information entropy of  $P$  concerning  $d$  is defined as follows:

$$H^\theta(P|d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{|R_P^{\theta x_i}|}. \tag{8}$$

**Lemma 3.16.**  $(X, B, d)$  is an HDIS. Let  $P \subseteq B$ ,  $\theta \in [0, 1]$  and  $D \in X/d$ . The following conclusion is valid:

$$(R_P^{\theta x} \cap D)(y) + (R_P^{\theta x} \cap (X - D))(y) = R_P^\theta(x, y) \ (\forall x, y \in X).$$

*Proof.*  $\forall x, y \in X$ , we have

$$(R_P^{\theta x} \cap D)(y) = R_P^\theta(x, y) \wedge D(y) = \begin{cases} R_P^\theta(x, y), & y \in D \\ 0, & y \notin D \end{cases}$$

$$(R_P^{\theta x} \cap (X - D))(y) = R_P^\theta(x, y) \wedge (X - D)(y) = \begin{cases} 0, & y \in D \\ R_P^\theta(x, y), & y \notin D \end{cases}$$

Thus,  $(R_P^{\theta x} \cap D)(y) + (R_P^{\theta x} \cap (X - D))(y) = R_P^\theta(x, y)$ .

**Proposition 3.17.**  $(X, B, d)$  is an HDIS. Let  $P \subseteq B$ ,  $\theta \in [0, 1]$  and  $D \in X/d$ . The following conclusion is valid:

$$|R_P^{\theta x} \cap D| + |R_P^{\theta x} \cap (X - D)| = |R_P^{\theta x}| \ (\forall x \in X).$$

*Proof.* According to Lemma 3.16,

$$\begin{aligned} |R_P^{\theta x} \cap D| + |R_P^{\theta x} \cap (X - D)| &= \sum_{i=1}^n (R_P^{\theta x} \cap D)(x_i) + \sum_{i=1}^n (R_P^{\theta x} \cap (X - D))(x_i) \\ &= \sum_{i=1}^n [(R_P^{\theta x} \cap D)(x_i) + (R_P^{\theta x} \cap (X - D))(x_i)] = \sum_{i=1}^n R_P^\theta(x, x_i) = |R_P^{\theta x}|. \end{aligned}$$

**Proposition 3. 18.**  $(X, B, d)$  is an HDIS.

1) If  $Q \subseteq P \subseteq B$ , then  $H^\theta(Q|d) \leq H^\theta(P|d)$ .

2) If  $0 \leq \theta_1 \leq \theta_2 \leq 1$ , then  $H^{\theta_1}(P|d) \leq H^{\theta_2}(P|d)$  ( $\forall P \subseteq B$ ).

*Proof.* 1) Let  $p_{ij}^{(1)} = |R_P^{\theta x_i} \cap D_j|$ ,  $p_{ij}^{(2)} = |R_P^{\theta x_i} \cap (X - D_j)|$ ,  $q_{ij}^{(1)} = |R_Q^{\theta x_i} \cap D_j|$ , and  $q_{ij}^{(2)} = |R_Q^{\theta x_i} \cap (X - D_j)|$ .

$R_Q^{\theta x_i} \subseteq R_P^{\theta x_i}$  according to Proposition 3.12.

Hence,  $0 \leq q_{ij}^{(1)} \leq p_{ij}^{(1)}$  and  $0 \leq q_{ij}^{(2)} \leq p_{ij}^{(2)}$ .

According to Proposition 3.17, we have  $p_{ij}^{(1)} + p_{ij}^{(2)} = |R_P^{\theta x_i}|$  and  $q_{ij}^{(1)} + q_{ij}^{(2)} = |R_Q^{\theta x_i}|$ .

Let  $f(x, y) = -x \log_2 \frac{x}{x+y}$  ( $x > 0, y > 0$ ).

$$\text{Then } H^\theta(P|d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{|R_P^{\theta x_i}|} = - \sum_{i=1}^n \sum_{j=1}^r \frac{p_{ij}^{(1)}}{n} \log_2 \frac{p_{ij}^{(1)}}{p_{ij}^{(1)} + p_{ij}^{(2)}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^r f(p_{ij}^{(1)}, p_{ij}^{(2)}) \text{ and}$$

$$H^\theta(Q|d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_\rho^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_\rho^{\theta x_i} \cap D_j|}{|R_\rho^{\theta x_i}|} = - \sum_{i=1}^n \sum_{j=1}^r \frac{q_{ij}^{(1)}}{n} \log_2 \frac{q_{ij}^{(1)}}{q_{ij}^{(1)} + q_{ij}^{(2)}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^r f(q_{ij}^{(1)}, q_{ij}^{(2)}).$$

Since the function  $f(x, y)$  exhibits monotonic increases in both  $x$  and  $y$ ,

$$f(q_{ij}^{(1)}, q_{ij}^{(2)}) \leq f(p_{ij}^{(1)}, q_{ij}^{(2)}) \leq f(p_{ij}^{(1)}, p_{ij}^{(2)}).$$

Hence,  $H^\theta(Q|d) \leq H^\theta(P|d)$ .

2) Let  $s_{ij}^{(1)} = |R_p^{\theta_1 x_i} \cap D_j|$ ,  $s_{ij}^{(2)} = |R_p^{\theta_1 x_i} \cap (X - D_j)|$ ,  $t_{ij}^{(1)} = |R_p^{\theta_2 x_i} \cap D_j|$ , and  $t_{ij}^{(2)} = |R_p^{\theta_2 x_i} \cap (X - D_j)|$ .

$R_p^{\theta_1 x_i} \subseteq R_p^{\theta_2 x_i}$  according to Proposition 3.12. Hence,  $0 \leq s_{ij}^{(1)} \leq t_{ij}^{(1)}$  and  $0 \leq s_{ij}^{(2)} \leq t_{ij}^{(2)}$ .

According to Proposition 3.17, we have  $s_{ij}^{(1)} + s_{ij}^{(2)} = |R_p^{\theta_1 x_i}|$  and  $t_{ij}^{(1)} + t_{ij}^{(2)} = |R_p^{\theta_2 x_i}|$ .

$$H^{\theta_1}(P|d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_p^{\theta_1 x_i} \cap D_j|}{n} \log_2 \frac{|R_p^{\theta_1 x_i} \cap D_j|}{|R_p^{\theta_1 x_i}|} = - \sum_{i=1}^n \sum_{j=1}^r \frac{s_{ij}^{(1)}}{n} \log_2 \frac{s_{ij}^{(1)}}{s_{ij}^{(1)} + s_{ij}^{(2)}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^r f(s_{ij}^{(1)}, s_{ij}^{(2)})$$

$$H^{\theta_2}(P|d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_p^{\theta_2 x_i} \cap D_j|}{n} \log_2 \frac{|R_p^{\theta_2 x_i} \cap D_j|}{|R_p^{\theta_2 x_i}|} = - \sum_{i=1}^n \sum_{j=1}^r \frac{t_{ij}^{(1)}}{n} \log_2 \frac{t_{ij}^{(1)}}{t_{ij}^{(1)} + t_{ij}^{(2)}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^r f(t_{ij}^{(1)}, t_{ij}^{(2)})$$

It follows from the monotonicity of  $f(x, y)$  that  $f(s_{ij}^{(1)}, s_{ij}^{(2)}) \leq f(t_{ij}^{(1)}, s_{ij}^{(2)}) \leq f(t_{ij}^{(1)}, t_{ij}^{(2)})$ .

Hence  $H^{\theta_1}(P|d) \leq H^{\theta_2}(P|d)$ .

**Definition 3.19.**  $(X, B, d)$  is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . The fuzzy joint information entropy of  $P$  and  $d$  is defined as follows:

$$H^\theta(P \cup d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_p^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_p^{\theta x_i} \cap D_j|}{n}. \tag{9}$$

**Lemma 3.20.**  $(X, B, d)$  is an HDIS. Let  $P \subseteq B$ ,  $\theta \in [0, 1]$ ,  $U \in 2^X$ , and  $V \in 2^X$ , then

$$\sum_{y \in U} R_p^\theta(x, y) + \sum_{y \in V} R_p^\theta(x, y) \geq \sum_{y \in U \cup V} R_p^\theta(x, y) \quad (\forall x \in X).$$

The equality holds as  $U \cap V = \Phi$ .

*Proof.* The conclusion is self-evident.

**Proposition 3.21.**  $(X, B, d)$  is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . Then

$$\sum_{j=1}^r |R_p^{\theta x} \cap D_j| = |R_p^{\theta x}| \quad (\forall x \in X).$$

*Proof.* Since  $D = \{D_1, D_2, \dots, D_r\}$  constitutes a partition of  $X$ ,

$$\begin{aligned} \sum_{j=1}^r |R_p^{\theta x} \cap D_j| &= \sum_{j=1}^r \sum_{y \in D_j} (R_p^\theta(x, y) \wedge D_j(y)) = \sum_{y \in X} (R_p^\theta(x, y) \wedge D(y)) \\ &= \sum_{y \in X} R_p^\theta(x, y) = |R_p^{\theta x}| \quad (\forall x \in X) \text{ by Lemma 3.20.} \end{aligned}$$

**Proposition 3.22.** *(X, B, d) is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . Then*

$$H^\theta(P|d) = H^\theta(P \cup d) - H^\theta(P).$$

*Proof.*  $\sum_{j=1}^r |R_P^{\theta x_i} \cap D_j| = |R_P^{\theta x_i}|$  according to Proposition 3.21.

$$\begin{aligned} H^\theta(P|d) &= - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{|R_P^{\theta x_i}|} \\ &= - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \left( \log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{n} - \log_2 \frac{|R_P^{\theta x_i}|}{n} \right) \\ &= - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{n} + \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i}|}{n} \\ &= H^\theta(P \cup d) + \sum_{i=1}^n \frac{|R_P^{\theta x_i}|}{n} \log_2 \frac{|R_P^{\theta x_i}|}{n} = H^\theta(P \cup d) - H^\theta(P). \end{aligned}$$

**Theorem 3.23.** *(X, B, d) is an HDIS. Let  $P \subseteq B$  and  $\theta \in [0, 1]$ . Then  $H^\theta(P|d) \geq 0$ .*

*Proof.*  $H^\theta(P) = - \sum_{i=1}^n \frac{|R_P^{\theta x_i}|}{n} \log_2 \frac{|R_P^{\theta x_i}|}{n}$  according to Definition 3.13.

$\sum_{j=1}^r |R_P^{\theta x_i} \cap D_j| = |R_P^{\theta x_i}|$  according to Proposition 3.21.

$$\text{Thus, } H^\theta(P) = - \sum_{i=1}^n \frac{\sum_{j=1}^r |R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i}|}{n} = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i}|}{n}.$$

$$H^\theta(P \cup d) = - \sum_{i=1}^n \sum_{j=1}^r \frac{|R_P^{\theta x_i} \cap D_j|}{n} \log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{n} \text{ according to Definition 3.19.}$$

Since  $\log_2 \frac{|R_P^{\theta x_i} \cap D_j|}{n} \leq \log_2 \frac{|R_P^{\theta x_i}|}{n}$ ,  $H^\theta(P) \leq H^\theta(P \cup d)$ .

$H^\theta(P|d) = H^\theta(P \cup d) - H^\theta(P)$  according to Proposition 3.22.

Hence,  $H^\theta(P|d) \geq 0$ .

### 3.4 An Attribute Reduction Algorithm Utilizing Fuzzy Conditional Information Entropy

**Algorithm 1:** An attribute reduction algorithm based on fuzzy conditional information entropy (IARFCIE)

---

**Input:** An HDIS  $(X, B, d)$  and  $\theta$ .  
**Output:** A reduct  $S$ .

- 1 Initialization:  $S \leftarrow B$ ;
- 2 Calculate  $X/d = \{D_1, D_2, \dots, D_r\}$ , where  $r$  denotes the number of distinct categories present in the dataset;
- 3 Calculate  $H^\theta(B|d)$  according to Definition 3.15;
- 4  $i = 0$ ;
- 5 **for** each attribute  $s \in S$  **do**
- 6     **if**  $H^\theta(S - s|d) = H^\theta(B|d)$  **then**
- 7          $S = S - s$ ;
- 8     **else**
- 9          $i = i + 1$ ;
- 10    **end**
- 11    **if**  $i = |S|$  **then**
- 12      Return  $S$ ;
- 13    **end**
- 14 **end**

---

**Definition 3.24.**  $(X, B, d)$  is an HDIS. Let  $A \subseteq B$  and  $\theta \in [0, 1]$ . If  $H^\theta(A|d) = H^\theta(B|d)$ ,  $A$  is called a coordination subset of  $B$ .

Let  $\text{coo}(B)$  denote the collection of all coordination subsets of  $B$ .

**Definition 3.25.**  $(X, B, d)$  is an HDIS. Let  $A \subseteq B$  and  $\theta \in [0, 1]$ . If  $A \in \text{coo}(B)$  and  $A - a \notin \text{coo}(B) (\forall a \in A)$ ,  $A$  is called a reduct of  $B$ .

Let  $\text{red}(B)$  denote the collection of all reducts of  $B$ .

In accordance with the aforementioned definitions and Proposition 3.18, Algorithm 1 for attribute reduction is hereby presented.

The time complexity of Algorithm 1 is shown in Table 3.

**Table 3:** The time complexity of IARFCIE

Step	Time complexity
2	$O( X )$
3	$O( X  B r)$
5–14	$O( X  B ^2r)$
Total	$O( X  B ^2r)$

## 4 Experimental Results and Discussions

### 4.1 Datasets

Across all experiments, we utilized twelve University of California Irvine (UCI) datasets with hybrid attributes to evaluate our approach. Table 4 offers comprehensive details on each dataset.

**Table 4:** Details of 12 UCI datasets

No.	Datesets	Abbreviation	Samples	Features	Classes
1	Abalone	Aba	4177	8	28
2	Automobile	Aut	205	25	6
3	Credit	Cre	690	15	2
4	Dermatology	Der	366	34	6
5	Hepatitis	Hep	155	19	2
6	Horse colic	HC	368	28	2
7	Processed cleveland	PC	303	13	5
8	Cylinder bands	CB	540	32	2
9	Insurance	Ins	9822	85	2
10	Mechanical analysis	MA	9254	7	6
11	German credit	GC	1000	20	2
12	Thyroid	Thy	7200	21	3

### 4.2 The Property of Monotonicity in Fuzzy Conditional Information Entropy

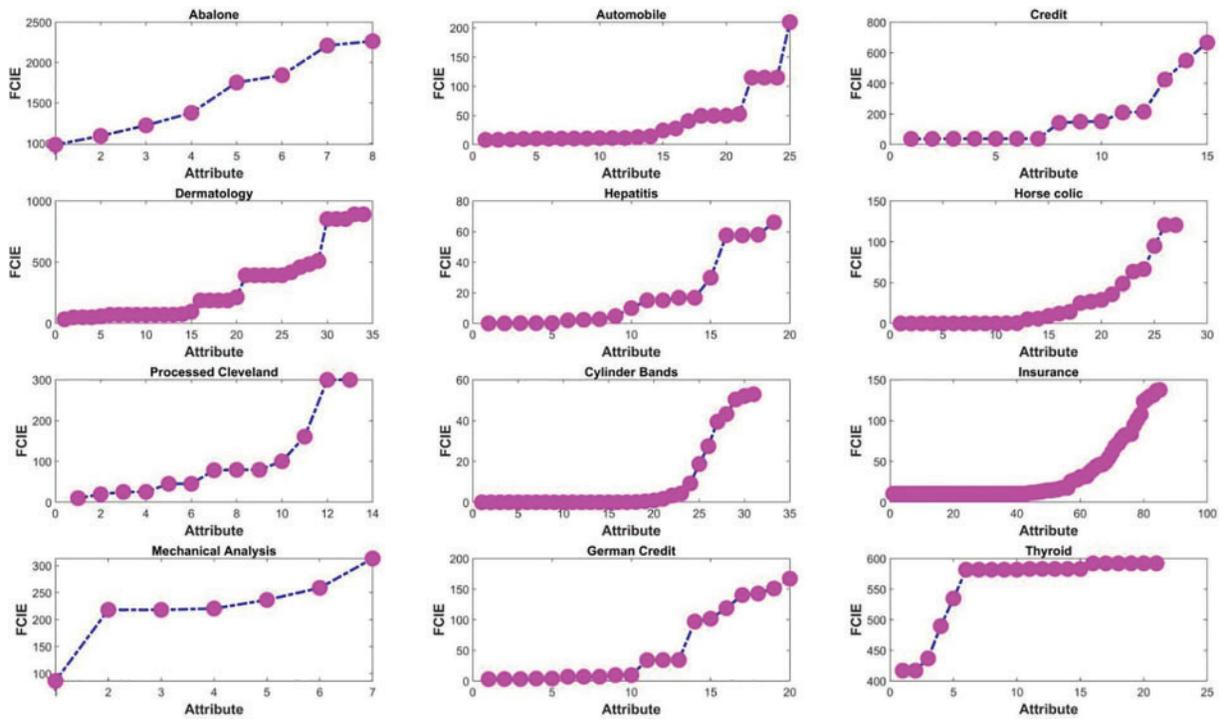
To verify the monotonicity of fuzzy conditional information entropy (FCIE), we calculate the fuzzy conditional information entropy of each dataset as the number of attributes increases.

Fig. 1 demonstrates that the fuzzy conditional information entropy consistently rises as the number of attributes increases, highlighting its monotonic behavior. The monotonicity displayed in Fig. 1 is a result of the parameter  $\theta$  set to 0.1. When other values are assigned to  $\theta$ , similar outcomes can also be observed.

### 4.3 The Influence of $\theta$ on Attribute Reduction and Classification Accuracy

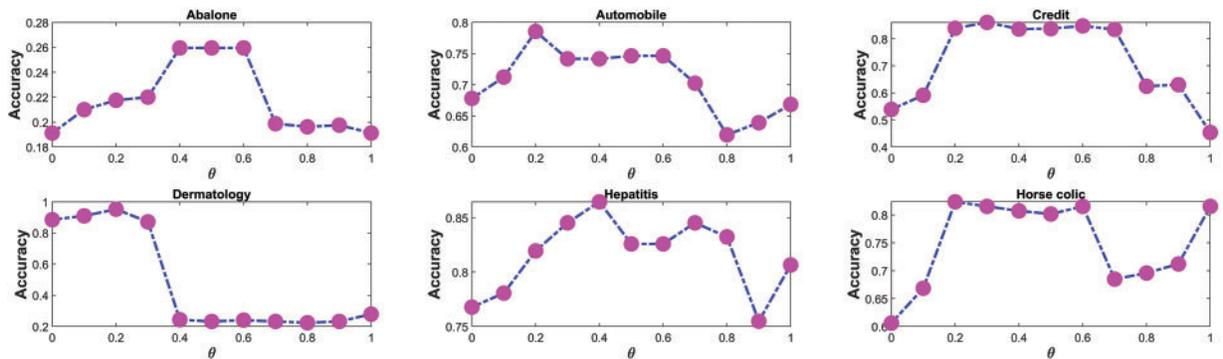
The parameter  $\theta$  of IARFCIE will affect the reduction results and classification accuracy. To determine the optimal parameter value, we can iterate through each value in  $\{0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1\}$  for  $\theta$  and select the one that yields the best performance.

In the experiment, a decision tree classifier was employed, and the obtained results were the mean of six iterations of 10-fold cross-validation. The complete dataset is evenly divided into 10 distinct subsets. Of these, 9 subsets are designated as training datasets, while the remaining subset serves as the test dataset. This rotation is repeated until the test dataset has undergone all 10 subsets, ensuring comprehensive evaluation. The number of epochs is 60.

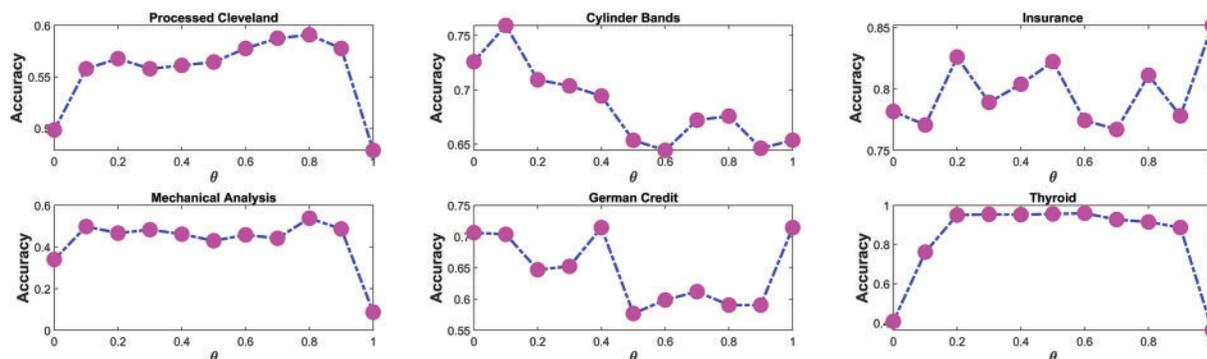


**Figure 1:** Monotonicity of fuzzy conditional information entropy

Fig. 2 shows the classification accuracy of the reduction results obtained by algorithm IARFCIE with different values of parameter  $\theta$ . By referring to Fig. 2, we can effectively identify the optimal parameter values for algorithm IARFCIE across the 12 datasets. Table 5 shows the optimal parameter values for 12 datasets.



**Figure 2:** (Continued)



**Figure 2:** Accuracy of classification for 12 datasets under different parameter settings

**Table 5:** The optimal parameter values of algorithm IARFCIE on 12 datasets

No.	Datasets	Abbreviation	$\theta$	Classification precision
1	Abalone	Aba	0.4, 0.5, 0.6	0.2595
2	Automobile	Aut	0.2	0.7805
3	Credit	Cre	0.3	0.8652
4	Dermatology	Der	0.2	0.9563
5	Hepatitis	Hep	0.2	0.8750
6	Horse colic	HC	0.1	0.8777
7	Processed cleveland	PC	0.1	0.6073
8	Cylinder bands	CB	0.1	0.7593
9	Insurance	Ins	0	0.9403
10	Mechanical analysis	MA	0.3	0.5567
11	German credit	GC	0.2	0.7440
12	Thyroid	Thy	0.1,02	0.9982

#### 4.4 Performance Comparison Results of IARFCIE and Other State-of-the-Art Attribute Reduction Algorithms

In this section, we present a comprehensive comparison of the proposed algorithm IARFCIE with nine state-of-the-art attribute reduction algorithms: The positive region backward deletion algorithm (PRBDA), the belief-based attribute reduction algorithm (BARA) [1], the cost-sensitive attribute-scale selection algorithm (CSASSA) [7], the entropy-approximate reduction algorithm (EARA) [11], the information-preserving reduction algorithm (IPRA) [26], the random forest algorithm (RFA) [27], the relief algorithm (RA) [28], the mutual information algorithm (MIA) [18], and the neighborhood rough set algorithm (NRSA) [2]. Tables 6 and 7 show the reduction results of 10 reduction algorithms on 12 datasets. Table 8 presents the classification accuracy achieved by 10 attribute reduction algorithms, as well as the original datasets, when utilizing decision tree (DT) classifier. The experimental results are obtained by averaging 10 runs of ten-fold cross-validation.

**Table 6:** The reduction results of five reduction algorithms on 12 datasets

Datasets	IARFCIE	PRBDA	EARA	IPRA	CSASSA
Aba	1, 3, 5, 8	1, 2, 3, 5, 6, 8	1, 3, 6, 8	5, 6, 7, 8	1
Aut	3, 6, 11, 12, 20	12, 5, 2, 1, 13, 10, 11, 3	22, 25	12, 11, 13, 1, 10, 5	2, 13
Cre	5, 6, 9, 10, 13	9, 8, 11, 15, 3, 4, 14	15, 14, 13, 12	5, 7, 8, 9, 11, 15	9, 11, 8, 15
Der	15, 5, 21, 28, 33, 6, 14, 2, 7	1, 2, 9, 14, 15, 16, 20, 22, 23, 24, 26, 28, 30	20, 22, 15, 21, 5, 28	5, 15, 26, 14, 21, 20, 22, 31, 28, 4, 7	1, 32, 17, 2, 3
Hep	9, 15, 17, 18	6, 11, 14, 17, 19	12, 11, 14	16, 1, 15	14, 17
HC	1, 13, 23, 25, 20, 7	1, 4, 11, 13, 15, 18, 20, 23, 25	1, 25, 13, 23	3, 12, 7, 15, 13	25, 1
PC	12, 3, 11, 9	8, 10, 5, 1, 12, 4, 13, 3	5, 8, 4	3, 8, 9, 10, 12, 13	12, 13
CB	8, 9, 24	1, 6, 7, 10, 11, 15, 17, 18, 19, 20, 21, 23, 24, 31	23, 22, 21, 14, 20	1, 2, 15, 20, 27, 29, 14	21
Ins	2	10, 16, 18, 21, 30, 32, 37, 42, 43, 44, 46, 47, 57, 59, 61, 64, 81, 82	59	1, 10, 35, 32, 5, 31, 7, 18, 15, 43	59
MA	3, 7	3, 7	3	4	7
GC	1, 3, 6	1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 20	5	1, 12, 20	5
Thy	3, 17, 19	17, 18, 21	3, 17, 21	17, 21, 19	17

**Table 7:** The reduction results of five reduction algorithms on 12 datasets

Datasets	RFA	NRSA	MIA	RA	BARA
Aba	1, 3	4	5, 6	5	1, 3, 6, 8
Aut	12, 1, 13, 10	25, 13, 1, 10	10, 1, 11, 12, 21, 13	18, 19, 8	10, 25, 22, 13, 21
Cre	11, 14, 8	15, 14, 2	9, 14	14, 2	1, 2, 5, 9, 11, 12, 14
Der	34, 15, 5	22, 27, 5, 15, 34, 4	33, 27, 29, 22, 12, 20, 6, 21, 25, 8	34, 31, 30	15, 5, 21, 28, 33, 6, 14, 11, 31, 26

(Continued)

**Table 7 (continued)**

Datasets	RFA	NRSA	MIA	RA	BARA
Hep	14, 12	16, 15, 18, 1	15, 17, 14, 18, 12	16, 15, 18, 1	8, 16, 18
HC	25, 1	3, 25	1, 25, 7	18, 12	3, 25, 18, 12
PC	12, 13	5, 1	5, 8	8, 5	3, 9, 11, 12
CB	21, 14, 19, 7, 22, 9, 1	22, 21, 14	25, 20, 5, 23, 2, 22	21, 17, 24, 19, 20	21, 19, 22, 17, 15, 1, 14, 31, 22, 19
Ins	1, 2	1, 59, 31, 18, 47	47, 61, 82, 68, 59, 1	6, 45, 12, 1, 52, 48	52, 61, 31, 47, 58, 5, 82
MA	7	3	7	3	7
GC	1	5, 2, 13	5, 1	5, 13, 2	5, 2, 13
Thy	17, 3, 21	7, 13, 21, 17, 19	17, 21, 18, 19	7, 13, 8	21, 13, 3, 17, 3

**Table 8:** The classification accuracy of 10 algorithms and raw datasets utilizing the DT classifier

Dataset	Raw dataset	IARFCIE	PRBDA	EARA	IPRA	CSASSA	RFA	NRSA	MIA	RA	BARA
Aba	0.2554	0.2595	0.2554	0.2554	0.2516	0.1754	0.2274	0.2516	0.2056	0.2495	0.2554
Aut	0.7268	0.7805	0.7317	0.4683	0.7122	0.5659	0.7610	0.6927	0.7366	0.5617	0.7472
Cre	0.8551	0.8652	0.8493	0.6681	0.8217	0.8159	0.7101	0.6565	0.8482	0.5812	0.8549
Der	0.9426	0.9563	0.8279	0.8115	0.8962	0.4672	0.5055	0.9344	0.7650	0.3661	0.9498
Hep	0.8000	0.8750	0.8065	0.7662	0.7548	0.8118	0.7857	0.8000	0.7792	0.8000	0.8599
HC	0.8207	0.8777	0.8587	0.8641	0.6793	0.8342	0.8342	0.8098	0.8342	0.7364	0.8491
PC	0.5809	0.6073	0.5380	0.5116	0.5512	0.5842	0.5842	0.5149	0.5380	0.5380	0.5886
CB	0.7333	0.7593	0.7148	0.7111	0.6778	0.5130	0.7278	0.7019	0.6907	0.7000	0.7275
Ins	0.9075	0.9403	0.9126	0.9403	0.9300	0.9403	0.9403	0.9406	0.9406	0.9402	0.9403
MA	0.5350	0.5567	0.5567	0.4812	0.4341	0.5451	0.4753	0.4196	0.4753	0.4196	0.4753
GC	0.7200	0.7440	0.7070	0.6000	0.7210	0.6000	0.7000	0.7110	0.6960	0.7110	0.7110
Thy	0.9960	0.9982	0.9779	0.9883	0.9729	0.9279	0.9783	0.9793	0.9808	0.9258	0.9802
Average	0.7394	0.7683	0.7280	0.6722	0.7002	0.6484	0.6858	0.7010	0.7075	0.6275	0.7449

#### 4.5 Discussions

Tables 6–8 demonstrate that IARFCIE outperforms all other algorithms and raw datasets across the 12 datasets when utilizing the DT classifier. Despite selecting more features than EARA, CSASSA, RFA, and RA, there is minimal difference in the number of selected features and IARFCIE significantly surpasses their accuracy.

CSASSA and RA selected very few features, but their accuracy was also very low, indicating that they suffered from underfitting. PRBDA selected a very large number of features, but its accuracy was not very high, indicating that it suffered from overfitting.

The superior performance of IARFCIE can be attributed to two factors: Firstly, it does not rely on Euclidean distance, instead utilizing a novel distance metric. As is widely recognized, Euclidean distance is unsuitable for assessing the differences between nominal attribute values. The novel distance

metric, which employs probability distributions to measure differences between nominal attribute values, aligns more closely with their inherent characteristics. Secondly, the algorithm introduces a new fuzzy relation that replaces similarity calculations based on distance with those derived from the number of attributes. This updated similarity measure effectively filters out a small number of outliers, thereby enhancing the robustness of the reduction algorithm.

## 5 Conclusions and Future Works

In this article, we introduce a novel difference metric that incorporates decision attributes. This metric offers a more accurate measurement of disparities between nominal attribute values. Subsequently, based on this new metric, we define a novel fuzzy relationship. This fuzzy relationship effectively filters out abnormal attribute values by utilizing the number of similar attributes to determine sample similarity. Furthermore, utilizing this new fuzzy relationship, we define a fuzzy conditional information entropy. An attribute reduction algorithm, formulated on the basis of fuzzy conditional information entropy, is then developed. Experimental results demonstrate that this attribute reduction algorithm not only exhibits a significant attribute reduction rate but also surpasses the original dataset as well as other attribute reduction algorithms in terms of average classification accuracy. Consequently, the novel metric and fuzzy relationship introduced in this article are effective in addressing the challenges associated with accurately measuring disparities between nominal attribute values and the sensitivity of attribute reduction algorithms towards abnormal attribute values. This study introduces a fresh perspective on reducing attributes in hybrid data by enhancing distance and fuzzy relationships. However, optimizing the parameters of the algorithm through grid search can significantly impact its efficiency. To address this issue, we aim to explore automatic parameter optimization methods as a future research direction.

**Acknowledgement:** The authors would like to thank the anonymous reviewers for their valuable comments.

**Funding Statement:** This work was supported by Anhui Province Natural Science Research Project of Colleges and Universities (2023AH040321) and Excellent Scientific Research and Innovation Team of Anhui Colleges (2022AH010098).

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Xiaoqin Ma; analysis and interpretation of results: Jun Wang, Wenchang Yu; draft manuscript preparation: Qinli Zhang. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** Not applicable.

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

## References

- [1] X. Chu *et al.*, “Multi-granularity dominance rough concept attribute reduction over hybrid information systems and its application in clinical decision-making,” *Inform. Sci.*, vol. 597, pp. 274–299, 2022. doi: [10.1016/j.ins.2022.03.048](https://doi.org/10.1016/j.ins.2022.03.048).

- [2] X. D. Fan, W. D. Zhao, C. Z. Wang, and Y. Huang, "Attribute reduction based on max-decision neighborhood rough set model," *Knowl-Based Syst.*, vol. 151, no. 1, pp. 16–23, 2018. doi: [10.1016/j.knosys.2018.03.015](https://doi.org/10.1016/j.knosys.2018.03.015).
- [3] W. H. Shu, W. B. Qin, and Y. H. Xie, "Incremental feature selection for dynamic hybrid data using neighborhood rough set," *Knowl-Based Syst.*, vol. 194, no. 22, pp. 28–39, 2020. doi: [10.1016/j.knosys.2020.105516](https://doi.org/10.1016/j.knosys.2020.105516).
- [4] J. Zhang, G. Q. Zhang, Z. W. Li, L. Qu, and C. Wen, "Feature selection in a neighborhood decision information system with application to single cell RNA data classification," *Appl. Soft Comput.*, vol. 113, no. 1, pp. 107876, 2021. doi: [10.1016/j.asoc.2021.107876](https://doi.org/10.1016/j.asoc.2021.107876).
- [5] P. Zhou, X. G. Hu, P. P. Li, and X. D. Wu, "Online streaming feature selection using adapted neighborhood rough set," *Inform. Sci.*, vol. 481, no. 5, pp. 258–279, 2019. doi: [10.1016/j.ins.2018.12.074](https://doi.org/10.1016/j.ins.2018.12.074).
- [6] S. Y. Xia, H. Zhang, W. H. Li, G. Y. Wang, E. Gien and Z. Z. Chen, "GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1231–1242, 2022. doi: [10.1109/TKDE.2020.2997039](https://doi.org/10.1109/TKDE.2020.2997039).
- [7] S. J. Liao, Y. D. Lin, J. J. Li, H. L. Li, and Y. H. Qian, "Attribute-scale selection for hybrid data with test cost constraint: The approach and uncertainty measures," *Int. J. Intell Syst.*, vol. 37, no. 6, pp. 3297–3333, 2022. doi: [10.1002/int.22678](https://doi.org/10.1002/int.22678).
- [8] Z. Yuan, H. Chen, T. Li, Z. Yu, B. Sang and C. Luo, "Unsupervised attribute reduction for mixed data based on fuzzy rough set," *Inform. Sci.*, vol. 572, no. 1, pp. 67–87, 2021. doi: [10.1016/j.ins.2021.04.083](https://doi.org/10.1016/j.ins.2021.04.083).
- [9] A. P. Zeng, T. R. Li, J. Hu, H. M. Chen, and C. Luo, "Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values," *Inform. Sci.*, vol. 378, no. 6, pp. 363–388, 2017. doi: [10.1016/j.ins.2016.07.056](https://doi.org/10.1016/j.ins.2016.07.056).
- [10] L. Yang, X. Y. Zhang, X. H. Xu, and B. B. Sang, "Multi-granulation rough sets and uncertainty measurement for multi-source fuzzy information system," *Int. J. Fuzzy Syst.*, vol. 21, no. 6, pp. 1919–1937, 2019. doi: [10.1007/s40815-019-00667-1](https://doi.org/10.1007/s40815-019-00667-1).
- [11] C. Z. Wang, Y. Wang, M. W. Shao, Y. H. Qian, and D. G. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 818–830, 2020. doi: [10.1109/TFUZZ.2019.2949765](https://doi.org/10.1109/TFUZZ.2019.2949765).
- [12] S. Singh, S. Shreevastava, T. Som, and G. Somani, "A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems," *Soft Comput.*, vol. 24, no. 6, pp. 4675–4691, 2020. doi: [10.1007/s00500-019-04228-4](https://doi.org/10.1007/s00500-019-04228-4).
- [13] P. Jain, A. K. Tiwari, and T. Som, "A fitting model based intuitionistic fuzzy rough feature selection," *Eng. Appl. Artif. Intell.*, vol. 89, pp. 1–13, 2020. doi: [10.1016/j.engappai.2019.103421](https://doi.org/10.1016/j.engappai.2019.103421).
- [14] Z. W. Li, Q. L. Zhang, S. P. Liu, Y. C. Peng, and L. L. Li, "Information fusion and attribute reduction for multi-source incomplete mixed data via conditional information entropy and D-S evidence theory," *Appl. Soft Comput.*, vol. 151, no. 11, pp. 111149, 2023. doi: [10.1016/j.asoc.2023.111149](https://doi.org/10.1016/j.asoc.2023.111149).
- [15] J. Vergara and P. Estevez, "A review of feature selection methods based on mutual information," *Neur. Comput. App.*, vol. 24, no. 1, pp. 175–186, 2014. doi: [10.1007/s00521-013-1368-0](https://doi.org/10.1007/s00521-013-1368-0).
- [16] Z. W. Li, Q. L. Zhang, P. Wang, Y. Song, and C. F. Wen, "Uncertainty measurement for a gene space based on class-consistent technology: An application in gene selection," *Appl. Intell.*, vol. 53, no. 2, pp. 5416–5436, 2022. doi: [10.1007/s10489-022-03657-3](https://doi.org/10.1007/s10489-022-03657-3).
- [17] Q. L. Zhang, Y. Y. Chen, G. Q. Zhang, Z. W. Li, L. J. Chen and C. F. Wen, "New uncertainty measurement for categorical data based on fuzzy information structures: An application in attribute reduction," *Inform. Sci.*, vol. 580, no. 5, pp. 541–577, 2021. doi: [10.1016/j.ins.2021.08.089](https://doi.org/10.1016/j.ins.2021.08.089).
- [18] X. Zhang, C. L. Mei, D. G. Chen, and J. H. Li, "Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy," *Pattern Recogn.*, vol. 56, pp. 1–15, 2016. doi: [10.1016/j.patcog.2016.02.013](https://doi.org/10.1016/j.patcog.2016.02.013).
- [19] B. B. Sang, H. M. Chen, L. Yang, T. R. Li, and W. H. Xu, "Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 6, pp. 1683–1697, 2021. doi: [10.1109/TFUZZ.2021.3064686](https://doi.org/10.1109/TFUZZ.2021.3064686).

- [20] Z. H. Huang and J. J. Li, "Discernibility measures for fuzzy  $\beta$  covering and their application," *IEEE Tansa. Cybernet.*, vol. 52, no. 9, pp. 9722–9735, 2022. doi: [10.1109/TCYB.2021.3054742](https://doi.org/10.1109/TCYB.2021.3054742).
- [21] M. Akram, H. S. Nawaz, and M. Deveci, "Attribute reduction and information granulation in Pythagorean fuzzy formal contexts," *Expert. Syst. Appl.*, vol. 222, no. 1167, pp. 119794, 2023. doi: [10.1016/j.eswa.2023.119794](https://doi.org/10.1016/j.eswa.2023.119794).
- [22] M. Akram, G. Ali, and J. C. R. Alcantud, "Attributes reduction algorithms for m-polar fuzzy relation decision systems," *Int. J. Approx. Reason.*, vol. 140, no. 3, pp. 232–254, 2022. doi: [10.1016/j.ijar.2021.10.005](https://doi.org/10.1016/j.ijar.2021.10.005).
- [23] L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, no. 3, pp. 338–353, 1965. doi: [10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- [24] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, no. 2–3, pp. 191–209, 2007. doi: [10.1080/03081079008935107](https://doi.org/10.1080/03081079008935107).
- [25] Z. Pawlak, "Rough sets," *Int. J. Inf. Comput. Sci.*, vol. 11, no. 5, pp. 341–356, 1982. doi: [10.1007/BF01001956](https://doi.org/10.1007/BF01001956).
- [26] Q. H. Hu, D. Yu, and Z. Xie, "Information-preserving hybrid data reduction based on fuzzy-rough techniques," *Pattern Recogn. Lett.*, vol. 27, no. 5, pp. 414–423, 2006. doi: [10.1016/j.patrec.2005.09.004](https://doi.org/10.1016/j.patrec.2005.09.004).
- [27] E. Sylvester *et al.*, "Applications of random forest feature selection for fine-scale genetic population assignment," *Evol. Appl.*, vol. 11, no. 2, pp. 153–165, 2018. doi: [10.1111/eva.12524](https://doi.org/10.1111/eva.12524).
- [28] R. J. Urbanowicz, M. Meeker, W. Lacava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Inform.*, vol. 85, no. 4, pp. 189–203, 2018. doi: [10.1016/j.jbi.2018.07.014](https://doi.org/10.1016/j.jbi.2018.07.014).