RESEARCH ARTICLE

New Measures of Uncertainty for Interval-Valued Data With Application to Attribute Reduction

LULU LI[ID]

School of Computer Science and Engineering, Yulin Normal University, Yulin, Guangxi 537000, China

e-mail: lululi100@126.com

ABSTRACT Uncertainty measurement (UM) gives a brand-new perspective on attribute reduction in an information system (IS). Interval-valued data is a kind of very vital data in rough set theory (RST). Rough set model based on tolerance relations can be considered to deal with interval-valued data. However, these kinds of tolerance relations have deficiencies when they are used in fuzzy rough computation. This paper studies new UM for an interval-valued information system (IVIS) and considers its attribute reduction. Firstly, a novel fuzzy symmetry relation on the object set of an IVIS is established based on "The similarity between information values that is fed back to the attribute set". Secondly, λ-information granules on the basis of a fuzzy symmetry relation are obtained. Then, four UMs for an IVIS are investigated. Next, numerical experiments and statistical tests are used to evaluate the performance of the proposed UMs. Moreover, attribute reduction in an IVIS is studied and the relevant algorithms are proposed. Finally, clustering analysis on the reduced IVIS is conducted. Experimental results indicate that the proposed algorithms are effective based on evaluation indicators of clustering performance. This paper provides a novel viewpoint for the establishment of fuzzy symmetry relation and attribute reduction algorithms.

INDEX TERMS RST, IVIS, UM, fuzzy symmetry relation, attribute reduction, cluster analysis.

I. INTRODUCTION

Rough set theory (RST), proposed by Pawlak [23], is a powerful tool to handle uncertainty. The completeness of RST depends on the data itself. In other words, it does not need to be attached to any additional information, so the results will become more objective and reliable. Nevertheless, two extended rough set models are raised for the sake of solving the disadvantage of too strict equivalence conditions. One is to introduce weak equivalence relations, such as to tolerance relations, similarity relations, dominance relations, or reflexive relations [34]; the other is to apply fuzzy set theory (FST) for RST [14].

An information system (IS) in the light of RST was raised by Pawlak. Uncertainty nearly exists in everywhere. Uncertainty measurement (UM) becomes a significant issue in a range of fields. The study of various UMs in ISs is helpful to understand the nature of information. UMs are able to be used for attribute reduction, rule acquisition, pattern recognition and clustering analysis. Some researchers have made some explorations in this respect and achieved a great deal of excellent research results. For instance, Düntsch et al. [11] investigated the measurement of decision rules on the basis of Shannon entropy in RST; Li et al. [20] gave UM methods based on fuzzy relation in an IS; Zeng et al. [42] researched UMs in a hybrid IS with the help of Gaussian kernel; Li et al. [17] provided a method to measure the uncertainty of a fully fuzzy IS by means of Gaussian kernel; Yang et al. [38] proposed UM methods for multi-source fuzzy IS.

Attribute reduction is a common technology in machine learning. Many datasets typically have a range of redundant attributes. Attribute reduction is aimed to eliminate extraneous and superfluous attributes from the original attributes of data and select the most effective attribute subset to eliminate the dimension of data. In recent years, many attribute reduction methods have emerged one after another. For example,

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei[ID].

Zeng et al. [40] provided a FST method for incremental attribute reduction in a hybrid IS; Cornelis et al. [5] gave a general definition for a fuzzy attribute reduction; Li et al. [16] studied UMs in view of description ability for attribute reduction in an IS; Chen et al. [4] proposed a attribute reduction method for heterogeneous data on the basis of the notion of fuzzy kernel alignment; Li et al. [18] advanced attribute reduction for heterogeneous data by means of information entropy; Wang et al. [31] put forward attribute reduction on account of local conditional entropy; Singh et al. [28] explored attribute reduction in an IVIS by employing a approach based on fuzzy similarity; Wang et al. [32] carried out attribute reduction by virtue of neighborhood self-information; Liu et al. [15] raised a familiar form of attribute reduction for an IS; Akram et al. [1] investigated parameter reduction under interval-valued m-polar fuzzy soft information; Ali et al. [2] researched attribute reduction in a bipolar fuzzy relation decision system.

FST was proposed by Zadeh [39] that describes the fuzziness in precise mathematical language. FST and RST, are two models to study inaccurate, uncertain and vague information. They have own advantages and features, and combine to form a model that is called fuzzy rough sets (FRSs) [7]. Wang et al. [29] put forward a fitting model for attribute reduction with FRSs; Chen et al. [6] raised a new attribute reduction algorithm on account of FRSs. Cornelis et al. [5] gave a extended model of attribute reduction in view of fuzzy tolerance relation within the context of FRSs. Yuan et al. [37] researched unsupervised attribute reduction for hybrid data by means of FRSs. Moreover, Wang et al. [34] proposed an integrated qualitative group decision-making method for assessing health-care waste treatment technologies based on linguistic terms with weakened hedges.

An interval-valued information system (IVIS) expresses an IS where its information values are interval-valued numbers. Based on the rich semantic explanations and flexibility, an IVIS have attracted attention of some scholars. Xie et al. [35] considered information structures of an IVIS and gave a new UM to measure uncertainty of the system. Dai et al. [8] studied UMs for an IVIS. Zhang et al. [41] proposed incremental updating of rough approximations in an IVIS under attribute generalization. Sakai et al. [27] presented a rule generation prototype system of IISs in Lipski that is able to process an IVIS.

Wang et al. [34] researched attribute reduction for categorical data based on FRSs. But they did not consider UMs and attribute reduction in an IVIS.

Usually, a tolerance relation on the object set of an IVIS is established according to "The similarity between information values that is fed back to the object set". Rough set model based on this tolerance relation is employed to dispose of interval-valued data. However, these kinds of tolerance relations have deficiencies when they are used in fuzzy rough computation. This paper introduces a novel fuzzy symmetry relation on the object set of an IVIS is established based on "The similarity between information values that is fed back to

the attribute set". Fuzzy rough set model based on this fuzzy symmetry relation is used to deal with interval-valued data. On this basis, this paper studies UM for interval-valued data with application to attribute reduction. The work process of this paper is displayed in FIGURE 1.

A. STRUCTURE AND ORGANIZATION

The specific arrangements of the article is structured as follows. In Section 2, some related concepts of fuzzy relations, interval-valued numbers and IVISs are raised. In Section 3, UMs of an IVIS are raised. In Section 4, some numerical experiments are designed to analyze the effectiveness of the proposed measures. In Section 5, an application for attribute reduction is given in an IVIS and clustering analysis on the reduced IVIS are doing. In Section 6, this paper is summarized.

II. PRELIMINARIES

In this article, O and AT signify two nonempty finite sets, $|X|$ means the cardinality of an ordinary set X, $F(O)$ indicates the family of all fuzzy subsets on O and $F(O \times O)$ typifies the family of all fuzzy relations on O. Put

$$O = \{o_1, o_2, \cdots, o_n\}, \quad AT = \{a_1, a_2, \cdots, a_m\} \quad \text{(II.1)}$$

A. FUZZY RELATIONS

A fuzzy subset S on O is a function $S : O \to [0, 1]$, and $S(x)$ means the membership degree of x to S. S is denoted as

$$S = \frac{S(o_1)}{o_1} + \frac{S(o_2)}{o_2} + \cdots + \frac{S(o_n)}{o_n} \quad \text{(II.2)}$$

$|S| = \sum_{i=1}^{n} S(o_i)$ shows the cardinality of S.

If $R \in O \times O$ is a fuzzy relation on O, and R may be denoted by $M(R) = (R(o_i, o_j))_{n \times n}$.

If $M(R)$ is a fuzzy identity matrix, and we denote by $R = \triangle$; if $R(o_i, o_j) = 1$, $i, j \leq n$ is a fuzzy universal relation, and we denote by $R = \omega$.

Let $R \in F(O \times O)$. $\forall o \in O$, define

$$R(o)(o') = R(o, o') \quad \text{(II.3)}$$

Then $R(o) \in F(O)$ can be treated as the information granule of o [24].

Obviously,

$$R(o) = \frac{R(o, o_1)}{o_1} + \frac{R(o, o_2)}{o_2} + \cdots + \frac{R(o, o_n)}{o_n} \quad \text{(II.4)}$$

Then

$$|R(o)| = \sum_{i=1}^{n} R(o, o_i).$$

B. INTERVAL-VALUED NUMBERS
Denote

$$[M] = \{s = [s^-, s^+] : s^-, s^+ \text{ are real numbers, } s^- \leq s^+\}$$
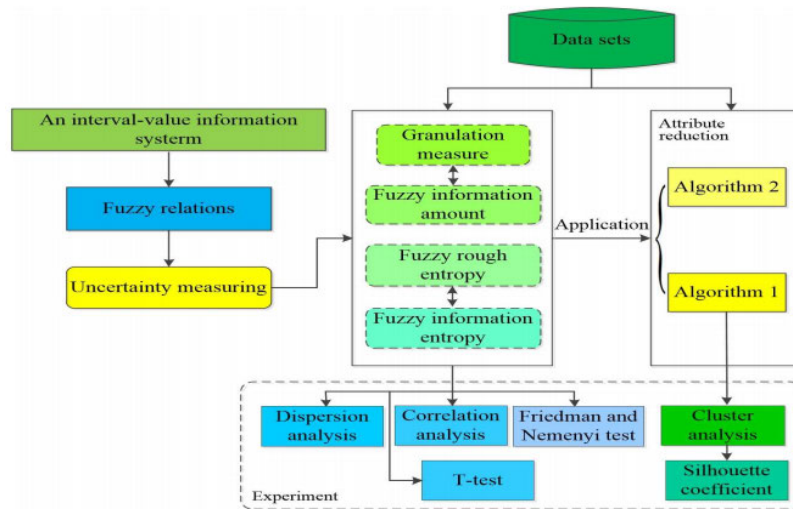
$$\text{(II.5)}$$

FIGURE 1. The work process of this paper.

Definition 1 [21], [22]: Given $s, t \in [M]$. Then the possible degree of s with respect to t is defined by

$$p(s, t) = min\{1, max\{\frac{s^+ - t^-}{(s^+ - s^-) + (t^+ - t^-)}, 0\}\} \quad \text{(II.6)}$$

Obviously,

a) $\forall s, t \in [M], 0 \leq p(s, t) \leq 1$;
b) $\forall s \in [M], p(s, s) = 0.5$;
c) $\forall s, t \in [M], p(s, t) + p(t, s) = 1$.

Definition 2 [9], [10]: Suppose $s, t \in [M]$. Then the similarity degree between m and t is defined as follows:

$$q(s, t) = 1 - |p(s, t) - p(t, s)| \quad \text{(II.7)}$$

Clearly,

(1) $\forall s, t \in [M], q(s, t) = q(t, s)$;
(2) $\forall s, t \in [M], 0 \leq q(s, t) \leq 1$;
(3) $\forall s, t \in [M], q(s, t) = 1 \Rightarrow s = t$.

Example 3: Pick $s = [6, 12]$ and $t = [3, 8]$. Then

$$p(s, t) = min\{1, max\{\frac{12 - 3}{(12 - 6) + (8 - 3)}, 0\}\} = \frac{9}{11},$$

$$p(t, s) = min\{1, max\{\frac{8 - 6}{(8 - 3) + (12 - 6)}, 0\}\} = \frac{2}{11},$$

$$q(s, t) = 1 - |p(s, t) - p(t, s)| = 1 - |\frac{9}{11} - \frac{2}{11}| = \frac{4}{11}.$$

C. AN IVIS

Definition 4: Let (O, AT) be an information system (IS). If $\forall a \in AT, \forall o \in O, a(o)$ is an interval-valued number, then (O, AT) is called an interval-valued information system (IVIS).

If $A \subseteq AT$, then (O, A) is known as the subsystem of (O, AT).

Example 5: TABLE 1 depicts an IVIS (O, AT) where $O = \{o_1, o_2, \cdots, o_{12}\}$ and $AT = \{a_1, a_2, \cdots, a_6\}$.

Let (O, AT) be an IVIS, $A \subseteq AT$ and $\lambda \in [0, 1]$. Define

$$T_A^\lambda = \{(o, o') \in O \times O : \forall a \in A, q(a(o), a(o')) \geq \lambda\} \quad \text{(II.8)}$$

Apparently, T_A^λ is a tolerance relation on O.

In T_A^λ, "$\forall a \in A, q(a(o), a(o')) \geq \lambda$" is fed back to the object set of an IVIS. Naturally, we may consider that "$\forall a \in A, q(a(o), a(o')) \geq \lambda$" is fed back to the attribute set of an IVIS. For this purpose, inspired by the paper [34], we introduce the following definition.

Definition 6: For an IVIS (O, AT), given $A \subseteq AT$ and $\lambda \in [0, 1]$. Then the fuzzy relation on O can be defined as

$$R_A^\lambda(o, o') = \frac{1}{m}|\{a \in A : q(a(o), a(o')) \geq \lambda\}|, \quad \forall o, o' \in O \quad \text{(II.9)}$$

Clearly, R_A^λ is a fuzzy symmetric relation on O.

$$\forall i, \quad R_A^\lambda(o_i) = \frac{R_A^\lambda(o_i, o_1)}{o_1} + \frac{R_A^\lambda(o_i, o_2)}{o_2} + \cdots + \frac{R_A^\lambda(o_i, o_n)}{o_n},$$

$$\forall i, \quad |R_A^\lambda(o_i)| = \sum_{i=1}^n R_A^\lambda(o_i, o_i).$$

For convenience, denote

$$A_{oo'}^\lambda = \{a \in A : q(a(o), a(o')) \geq \lambda\} \quad \text{(II.10)}$$

Then

$$R_A^\lambda(o, o') = \frac{1}{m}|A_{oo'}^\lambda| \quad \text{(II.11)}$$

Obviously,

$$R_A^\lambda(o, o) = \frac{|A|}{m} \text{ and } \forall o \in O, |R_A^\lambda(o)| \geq \frac{|A|}{m}$$

TABLE 1. An IVIS.

	a_1	a_2	a_3	a_4	a_5	a_6
o_1	[69.71,73.57]	[14.82,15.91]	[53.05,55.23]	[43.67,47.51]	[65.38,69.33]	[52.97,55.61]
o_2	[70.19,71.94]	[12.43,12.96]	[52.81,53.56]	[44.08,46.06]	[53.35,54.94]	[45.14,47.53]
o_3	[69.54,72.85]	[22.28,22.83]	[60.05,62.98]	[38.85,42.89]	[67.24,69.91]	[55.11,59.11]
o_4	[74.18,76.26]	[18.75,19.06]	[52.22,53.53]	[30.39,34.17]	[69.18,70.25]	[48.81,49.95]
o_5	[63.13,67.34]	[16.75,18.78]	[57.83,58.05]	[36.62,38.75]	[60.29,60.85]	[45.74,48.23]
o_6	[70.64,74.84]	[18.67,20.91]	[57.89,61.67]	[38.49,39.57]	[67.69,68.37]	[52.02,56.52]
o_7	[63.54,64.83]	[17.54,19.37]	[62.19,63.41]	[35.41,39.36]	[61.76,65.15]	[53.73,56.61]
o_8	[61.79,64.86]	[18.77,22.59]	[57.73,59.52]	[33.94,38.69]	[53.11,55.57]	[56.97,60.93]
o_9	[69.11,72.02]	[17.53,20.67]	[56.11,59.55]	[46.43,48.07]	[54.24,54.95]	[30.01,33.71]
o_{10}	[66.75,69.46]	[19.65,23.51]	[59.73,61.52]	[38.16,41.95]	[58.14,60.61]	[53.37,56.83]
o_{11}	[66.88,71.23]	[17.27,21.93]	[60.64,64.32]	[47.76,49.95]	[64.97,65.71]	[46.54,47.77]
o_{12}	[69.93,71.25]	[22.92,27.79]	[57.76,59.73]	[37.82,41.99]	[66.51,66.79]	[56.73,60.06]

Theorem 7: For an IVIS (O, AT), if $A \subseteq B \subseteq AT$, then $\forall \lambda \in [0, 1], R_A^\lambda \subseteq R_B^\lambda$.

Proof: By Definition 6,

$$R_A^\lambda(o, o') = \frac{1}{m}|A_{oo'}^\lambda|, \quad R_B^\lambda(o, o') = \frac{1}{m}|B_{oo'}^\lambda|.$$

Since $A \subseteq B \subseteq AT$, we have $\forall o, o' \in O, A_{oo'}^\lambda \subseteq B_{oo'}^\lambda$.
So $\forall o, o' \in O, R_A^\lambda(o, o') \leq R_B^\lambda(o, o')$.
Thus, $R_A^\lambda \subseteq R_B^\lambda$. ∎

Theorem 8: For an IVIS (O, AT), if $0 \leq \lambda_1 \leq \lambda_2 \leq 1$, then $\forall A \subseteq AT, R_A^{\lambda_2} \subseteq R_A^{\lambda_1}$.

Proof: By Definition 6,

$$R_A^{\lambda_1}(o, o') = \frac{1}{m}|A_{oo'}^{\lambda_1}|, \quad R_A^{\lambda_2}(o, o') = \frac{1}{m}|A_{oo'}^{\lambda_2}|.$$

Since $0 \leq \lambda_1 \leq \lambda_2 \leq 1$, we have $\forall o, o' \in O, A_{oo'}^{\lambda_2} \subseteq A_{oo'}^{\lambda_1}$.
So $\forall o, o' \in O, R_A^{\lambda_2}(o, o') \leq R_A^{\lambda_1}(o, o')$.
Thus, $R_A^{\lambda_2} \subseteq R_A^{\lambda_1}$. ∎

III. MEASURING UNCERTAINTY OF AN IVIS

This part puts forward some tools for measuring uncertainty of an IVIS.

A. GRANULATION MEASURE FOR AN IVIS

Definition 9: For an IVIS (O, AT), given $A \subseteq AT$. Then fuzzy information granulation of (O, A) is specified as

$$G^\lambda(A) = \frac{1}{n^2} \sum_{i=1}^{n} |R_A^\lambda(o_i)| \qquad \text{(III.1)}$$

Example 10: Table 1 illustrates an an IVIS. Pick $A = \{a_1, a_2, a_3\}$ and $\lambda = 0.2$, Then

$$G^\lambda(A) \approx 0.2384.$$

Proposition 11: For an IVIS (O, AT), given $A \subseteq AT$. Then

$$\frac{|A|}{nm} \leq G^\lambda(A) \leq 1.$$

Furthermore, if $R_A^\lambda = \triangle$, $G^\lambda(A) = \frac{1}{n}$; if $R_A^\lambda = \omega$, then $G^\lambda(A) = 1$.

Proof: Since $\forall i, \frac{|A|}{m} \leq |R_A^\lambda(o_i)| \leq n$, we have

$$n\frac{|A|}{m} \leq \sum_{i=1}^{n} |R_A^\lambda(o_i)| \leq n^2.$$

By Definition 9,

$$\frac{|A|}{nm} \leq G^\lambda(A) \leq 1.$$

Suppose $R_A^\lambda = \triangle$. Then $\forall i, |R_A^\lambda(o_i)| = 1, G^\lambda(A) = \frac{1}{n}$.
Suppose $R_A^\lambda = \omega$. Then $\forall i, |R_A^\lambda(o_i)| = n, G^\lambda(A) = 1$. ∎

Proposition 12: For an IVIS (O, AT), if $A \subseteq B \subseteq AT$, then $G^\lambda(A) \leq G^\lambda(B)$.

Proof: Since $A \subseteq B \subseteq AT$, by Theorem 7, we have $R_A^\lambda \subseteq R_B^\lambda$.
Then $\forall i, R_A^\lambda(o_i) \subseteq R_B^\lambda(o_i)$. Thus

$$\forall i, \quad |R_A^\lambda(o_i)| \leq |R_B^\lambda(o_i)|.$$

By Definition 9,

$$G^\lambda(A) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n}|R_A^\lambda(o_i)|, \quad G^\lambda(B) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n}|R_B^\lambda(o_i)|.$$

Hence $G^\lambda(A) \leq G^\lambda(B)$. ∎

B. ENTROPY MEASUREMENT FOR AN IVIS

Definition 13: For an IVIS (O, AT), given $A \subseteq AT$. Then fuzzy information entropy of (O, A) is defined as

$$H^\lambda(A) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{|R_A^\lambda(o_i)|}{n} \qquad \text{(III.2)}$$

Example 14 (Continued From Example 10): We have

$$H^\lambda(A) \approx 2.1002.$$

Proposition 15: For an IVIS (O, AT), if $A \subseteq B \subseteq AT$, then $H^\lambda(B) \leq H^\lambda(A)$.

Proof: Since $A \subseteq B \subseteq AT$, by Theorem 7, we have $R_A^\lambda \subseteq R_B^\lambda$.
Then $\forall i, R_A^\lambda(o_i) \subseteq R_B^\lambda(o_i)$. So

$$\forall i, \quad |R_A^\lambda(o_i)| \leq |R_B^\lambda(o_i)|.$$

Thus $\forall i$,

$$-\log_2 \frac{|R_A^\lambda(o_i)|}{n} = \log_2 \frac{n}{|R_A^\lambda(o_i)|}$$

$$\geq \log_2 \frac{n}{|R_B^\lambda(o_i)|} = -\log_2 \frac{|R_B^\lambda(o_i)|}{n},$$

Consequently, $H^\lambda(B) < H^\lambda(A)$. ∎

Definition 16: For an IVIS (O, AT), given $A \subseteq AT$. Then fuzzy rough entropy of (O, A) is deemed as

$$E_r^\lambda(A) = -\sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{|R_A^\lambda(o_i)|} \qquad (III.3)$$

Example 17 (Continued From Example 10): We have

$$E_r^\lambda(A) \approx 1.4847.$$

Proposition 18: For an IVIS (O, AT), given $A \subseteq AT$. Then

$$\log_2 \frac{|A|}{m} \leq E_r^\lambda(A) \leq \log_2 n \qquad (III.4)$$

Moreover, if $R_A^\lambda = \triangle$, *then* $E_r^\lambda(A) = 0$; *if* $R_A^\lambda = \omega$, *then* $E_r^\lambda(A) = \log_2 n$.

Proof: Since $\forall i$, $\frac{|A|}{m} \leq |R_A^\lambda(o_i)| \leq n$, we have

$$\log_2 \frac{|A|}{m} \leq -\log_2 \frac{1}{|R_A^\lambda(o_i)|} = \log_2 |R_A^\lambda(o_i)| \leq \log_2 n.$$

By Definition 16, we obtain that

$$\log_2 \frac{|A|}{m} \leq E_r^\lambda(A) \leq \log_2 n.$$

Suppose $R_A^\lambda = \triangle$. Then $\forall i$, $|R_A^\lambda(o_i)| = 1$. Thus $E_r^\lambda(A) = 0$.
Suppose $R_A^\lambda = \omega$. Then $\forall i$, $|R_A^\lambda(o_i)| = n$. Thus $E_r^\lambda(A) = \log_2 n$. ∎

Proposition 19: For an IVIS (O, AT), if $A \subseteq B \subseteq AT$, then $E_r^\lambda(A) \leq E_r^\lambda(B)$.

Proof: Since $A \subseteq B \subseteq AT$, by Theorem 7, we have $R_A^\lambda \subseteq R_B^\lambda$.
Then $\forall i$, $R_A^\lambda(o_i) \subseteq R_B^\lambda(o_i)$. So

$$\forall i, \quad |R_A^\lambda(o_i)| \leq |R_B^\lambda(o_i)|.$$

Thus $\forall i$,

$$-\log_2 \frac{1}{|R_A^\lambda(o_i)|} = \log_2 |R_A^\lambda(o_i)|$$

$$\leq \log_2 |R_B^\lambda(o_i)| = -\log_2 \frac{1}{|R_B^\lambda(o_i)|}$$

Thus, $E_r^\lambda(A) < E_r^\lambda(B)$. ∎

Theorem 20: For an IVIS (O, AT), given $A \subseteq AT$. Then

$$E_r^\lambda(A) + H^\lambda(A) = \log_2 n \qquad (III.5)$$

Proof:

$$E_r^\lambda(A) + H^\lambda(A) = -\frac{1}{n} \sum_{i=1}^n \left(\log_2 \frac{1}{|R_A^\lambda(o_i)|} + \log_2 \frac{|R_A^\lambda(o_i)|}{n} \right)$$

$$= -\frac{1}{n} \sum_{i=1}^n \log_2 \frac{1}{n} = \log_2 n.$$

∎

Corollary 21: For an IVIS (O, AT), given $A \subseteq AT$. Then

$$0 \leq H^\lambda(A) \leq \log_2 \frac{nm}{|A|} \qquad (III.6)$$

Moeover, if $R_A^\lambda = \triangle$, *then* $H^\lambda(A) = \log_2 n$; *if* $R_A^\lambda = \omega$, *then* $H^\lambda(A) = 0$.

Proof: By Proposition 18, $\log_2 \frac{|A|}{m} \leq E_r^\lambda(A) \leq \log_2 n$.
By Theorem 20, $H^\lambda(A) = \log_2 n - E_r^\lambda(A)$.
Consequently, $0 \leq H^\lambda(A) \leq \log_2 n - \log_2 \frac{|A|}{m} = \log_2 \frac{nm}{|A|}$.
Suppose $R_A^\lambda = \triangle$. Then $\forall i$, $|R_A^\lambda(o_i)| = 1$. Thus $H^\lambda(A) = \log_2 n$.
Suppose $R_A^\lambda = \omega$. Then $\forall i$, $|R_A^\lambda(o_i)| = n$. Thus $H^\lambda(A) = 0$. ∎

C. FUZZY INFORMATION AMOUNT IN AN IVIS

Definition 22: For an IVIS (O, AT), given $A \subseteq AT$. Then fuzzy information amount of (O, A) is regarded as

$$E^\lambda(A) = \sum_{i=1}^n \frac{1}{n} \left(1 - \frac{|R_A^\lambda(o_i)|}{n}\right) \qquad (III.7)$$

Example 23 (Continued From Example 10): We have

$$E^\lambda(A) \approx 0.7616.$$

Proposition 24: For an IVIS (O, AT), if $A \subseteq B \subseteq AT$, then $E^\lambda(B) \leq E^\lambda(A)$.

Proof: Since $A \subseteq B \subseteq AT$, by Theorem 7, we have $R_A^\lambda \subseteq R_B^\lambda$.
Then $\forall i$, $R_A^\lambda(o_i) \subseteq R_B^\lambda(o_i)$. So

$$\forall i, \quad |R_A^\lambda(o_i)| \leq |R_B^\lambda(o_i)|.$$

Thus $\forall i$, $1 \leq |R_A^\lambda(o_i)| \leq |R_B^\lambda(o_i)|$.
Hence $E^\lambda(B) < E^\lambda(A)$. ∎

Theorem 25: For an IVIS (O, AT), given $A \subseteq AT$. Then

$$G^\lambda(A) + E^\lambda(A) = 1 \qquad (III.8)$$

Proof:

$$G^\lambda(A) + E^\lambda(A) = \frac{1}{n^2} \sum_{i=1}^n [|R_A^\lambda(o_i)| + (n - |R_A^\lambda(o_i)|)] = 1.$$

∎

Corollary 26: For an IVIS (O, AT), given $A \subseteq AT$. Then

$$0 \leq E^\lambda(A) \leq 1 - \frac{|A|}{nm} \qquad (III.9)$$

Furthermore, if $R_A^\lambda = \triangle$, *then* $E^\lambda(A) = 1 - \frac{1}{n}$; *if* $R_A^\lambda = \omega$, *then* $E^\lambda(A) = 0$.

Proof: By Proposition 11, $\frac{|A|}{nm} \leq G^\lambda(A) \leq 1$.
By Theorem 25, $E^\lambda(A) = 1 - G^\lambda(A)$.
Thus

$$0 \leq E^\lambda(A) \leq 1 - \frac{|A|}{nm}.$$

Suppose $R_A^\lambda = \triangle$. Then $\forall i$, $|R_A^\lambda(o_i)| = 1$, $E^\lambda(A) = 1 - \frac{1}{n}$.
Suppose $R_A^\lambda = \omega$. Then $\forall i$, $|R_A^\lambda(o_i)| = n$, $E^\lambda(A) = 0$. ∎

IV. EXPERIMENTAL EVALUATION

This part designs some experiments and does effectiveness analysis to the proposed measures.

TABLE 2. Six data sets from UCI.

No.	Date sets	Abbr.	Objects	Attributes
1	Iris	Ir	150	4
2	Ecoli	Ec	336	7
3	Parkinsons	Pa	197	23
4	Seeds	Se	210	7
5	Sona	So	208	30
6	Wdbc	Wd	369	30

A. DATA SETS AND EXPERIMENTAL CONTENTS

Six data sets from UCI (Machine Learning Repository) are picked (see TABLE 2).

The above data sets come from UCI. They are all real number type. They can change to interval-valued data. Then, the interval-valued number $a'(x)$ converted from the information value of the object o_i under the attribute a_i can be obtained by formula $a'(x) = [a(x) - \xi\sigma, a(x) + \xi\sigma]$, where σ is the standard deviation. Usually, we pick $\xi = 5$.

For the dataset Ir, denote $L_i = \{a_1, \cdots, a_i\}$ $(i = 1, \cdots, 4)$. Define

$$X_{G^\lambda}(Ir) = \{G^\lambda(L_1), \cdots, G^\lambda(L_4)\},$$
$$X_{E^\lambda}(Ir) = \{E^\lambda(L_1), \cdots, E^\lambda(L_4)\},$$
$$X_{E_r^\lambda}(Ir) = \{E_r^\lambda(L_1), \cdots, E_r^\lambda(L_4)\},$$
$$X_{H^\lambda}(Ir) = = \{H^\lambda(L_1), \cdots, H^\lambda(L_4)\}.$$

For the dataset Ec, denote $M_i = \{a_1, \cdots, a_i\}$ $(i = 1, \cdots, 7)$. Define

$$X_{G^\lambda}(Ec) = \{G^\lambda(M_1), \cdots, G^\lambda(M_7)\},$$
$$X_{E^\lambda}(Ec) = \{E^\lambda(M_1), \cdots, E^\lambda(M_7)\},$$
$$X_{E_r^\lambda}(Ec) = \{E_r^\lambda(M_1), \cdots, E_r^\lambda(M_7)\},$$
$$X_{H^\lambda}(Ec) = = \{H^\lambda(M_1), \cdots, H^\lambda(M_7)\}.$$

For the dataset Pa, denote $N_i = \{a_1, \cdots, a_i\}$ $(i = 1, \cdots, 23)$. Define

$$X_{G^\lambda}(Pa) = \{G^\lambda(N_1), \cdots, G^\lambda(N_{23})\},$$
$$X_{E^\lambda}(Pa) = \{E^\lambda(N_1), \cdots, E^\lambda(N_{23})\},$$
$$X_{E_r^\lambda}(Pa) = \{E_r^\lambda(N_1), \cdots, E_r^\lambda(N_{23})\},$$
$$X_{H^\lambda}(Pa) = \{H^\lambda(N_1), \cdots, H^\lambda(N_{23})\}.$$

For the dataset Se, denote $O_i = \{a_1, \cdots, a_i\}$ $(i = 1, \cdots, 7)$. Define

$$X_{G^\lambda}(Se) = \{G^\lambda(O_1), \cdots, G^\lambda(O_7)\},$$
$$X_{E^\lambda}(Se) = \{E^\lambda(O_1), \cdots, E^\lambda(O_7)\},$$
$$X_{E_r^\lambda}(Se) = \{E_r^\lambda(O_1), \cdots, E_r^\lambda(O_7)\},$$
$$X_{H^\lambda}(Se) = \{H^\lambda(O_1), \cdots, H^\lambda(O_7)\}.$$

For the dataset So, denote $A_i = \{a_1, \cdots, a_{2\times i}\}$ $(i = 1, \cdots, 15)$. Define

$$X_{G^\lambda}(So) = \{G^\lambda(A_1), \cdots, G^\lambda(A_{15})\},$$
$$X_{E^\lambda}(So) = \{E^\lambda(A_1), \cdots, E^\lambda(A_{15})\},$$
$$X_{E_r^\lambda}(So) = \{E_r^\lambda(A_1), \cdots, E_r^\lambda(A_{15})\},$$
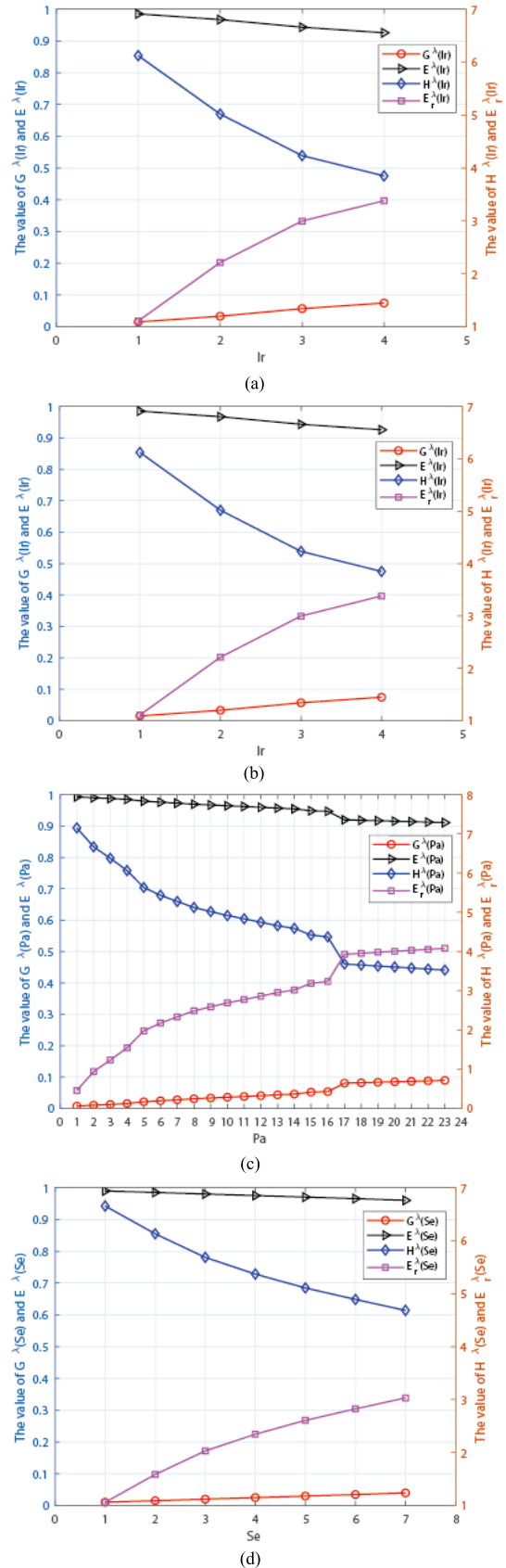$$X_{H^\lambda}(So) = \{H^\lambda(A_1), \cdots, H^\lambda(A_{15})\}.$$

FIGURE 2. *pcc.*

TABLE 3. The corresponding correlation between *S* and *T*.

$pcc(S,T)$	Correlation between X and T	Abbreviation
$pcc(S,T) = -1$	Completely negative correlation	CNC
$-1 \leq pcc(S,T) < -0.7$	Height negative correlation	HNC
$-0.7 \leq pcc(S,T) < -0.4$	Moderate negative correlation	MNC
$-0.4 < pcc(S,T) < 0$	Low negative correlation	LNC
$pcc(S,T) = 0$	No correlation	NC
$0 < pcc(S,T) < 0.4$	Low positive correlation	LPC
$0.4 \leq pcc(S,T) < 0.7$	Moderate positive correlation	MPC
$0.7 \leq pcc(S,T) < 1$	Height positive correlation	HPC
$pcc(S,T) = 1$	Completely positive correlation	CPC

(e)

(f)

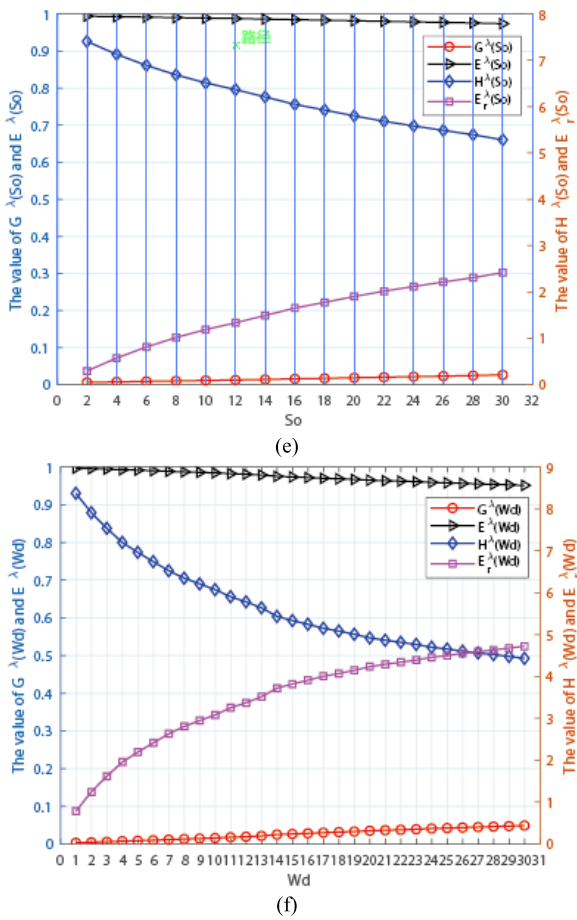**FIGURE 2.** (Continued.) pcc.

For the dataset Wd, denote $Q_i = \{a_1, \cdots, a_i\}$ ($i = 1, \cdots, 30$). Define

$$X_{G^\lambda}(Wd) = \{G^\lambda(Q_1), \cdots, G^\lambda(Q_{30})\},$$
$$X_{E^\lambda}(Wd) = \{E^\lambda(Q_1), \cdots, E^\lambda(Q_{30})\},$$
$$X_{E_r^\lambda}(Wd) = \{E_r^\lambda(Q_1), \cdots, E_r^\lambda(Q_{30})\},$$
$$X_{H^\lambda}(Wd) = \{H^\lambda(Q_1), \cdots, H^\lambda(Q_{30})\}.$$

B. EXPERIMENTAL RESULTS

From FIGUREs $2a - 2f$, we obtain the following experimental results.

G^λ, E_r^λ are both monotonically increases and E^λ, H^λ are both monotonically decreasing as the cardinality of attribute subset increases. Thus, G^λ, E^λ, E_r^λ and H^λ are applied to measure the uncertainty of an IVIS.

C. DISPERSION ANALYSIS

Define

$$CV(S) = \frac{\sigma(S)}{\bar{s}} \qquad \text{(IV.1)}$$

where

$$S = \{s_1, \cdots, s_n\}, \quad \bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i, \ \sigma(S) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_i - \bar{s})^2}.$$

$CV(S)$ is called the standard deviation coefficient of S. It is used for analyzing the effectiveness of the proposed measures

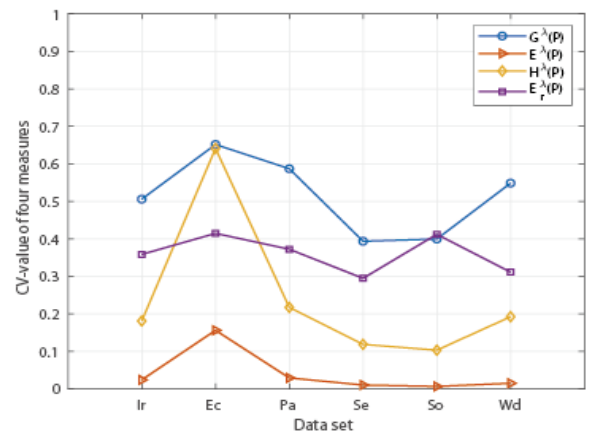Continue the above experiment, CV-values of four measure sets are shown in FIGURE 3.

FIGURE 3. The work process of this paper.

From the FIGURE 3, we can draw a conclusion that the CV-values of G^λ and E_r^λ are much higher than those of E^λ and H^λ. For these six datasets, the CV-value of E^λ is obviously the smallest and indicates that the dispersion degree of E_r^λ is minimum, that is to say, E^λ has much better performance for measuring uncertainty of an IVIS.

TABLE 4. *pcc*-values of sixteen pairs of four measure sets on Ir.

pcc	$X_{G^\lambda}(Ir)$	$X_{E^\lambda}(Ir)$	$X_{H^\lambda}(Ir)$	$X_{E_r^\lambda}(Ir)$
$X_{G^\lambda}(Ir)$	1			
$X_{E^\lambda}(Ir)$	-1	1		
$X_{H^\lambda}(Ir)$	-0.9779	0.9779	1	
$X_{E_r^\lambda}(Ir)$	0.9779	-0.9779	-1	1

TABLE 5. *pcc*-values of sixteen pairs of four measure sets on Ec.

pcc	$X_{G^\lambda}(Ec)$	$X_{E^\lambda}(Ec)$	$X_{H^\lambda}(Ec)$	$X_{E_r^\lambda}(Ec)$
$X_{G^\lambda}(Ec)$	1			
$X_{E^\lambda}(Ec)$	-1	1		
$X_{H^\lambda}(Ec)$	-0.9692	0.9692	1	
$X_{E_r^\lambda}(Ec)$	0.9692	-0.9692	-1	1

TABLE 6. *pcc*-values of sixteen pairs of four measure sets on Pa.

pcc	$X_{G^\lambda}(Pa)$	$X_{E^\lambda}(Pa)$	$X_{H^\lambda}(Pa)$	$X_{E_r^\lambda}(Pa)$
$X_{G^\lambda}(Pa)$	1			
$X_{E^\lambda}(Pa)$	-1	1		
$X_{H^\lambda}(Pa)$	-0.9547	0.9547	1	
$X_{E_r^\lambda}(Pa)$	0.9547	-0.9547	-1	1

TABLE 7. *pcc*-values of sixteen pairs of four measure sets on Se.

pcc	$X_{G^\lambda}(Se)$	$X_{E^\lambda}(Se)$	$X_{H^\lambda}(Se)$	$X_{E_r^\lambda}(Se)$
$X_{G^\lambda}(Se)$	1			
$X_{E^\lambda}(Se)$	-1	1		
$X_{H^\lambda}(Se)$	-0.9815	0.9815	1	
$X_{E_r^\lambda}(Se)$	0.9815	-0.9815	-1	1

TABLE 8. *pcc*-values of sixteen pairs of four measure sets on So.

pcc	$X_{G^\lambda}(So)$	$X_{E^\lambda}(So)$	$X_{H^\lambda}(So)$	$X_{E_r^\lambda}(So)$
$X_{G^\lambda}(So)$	1			
$X_{E^\lambda}(So)$	-1	1		
$X_{H^\lambda}(So)$	-0.9813	0.9813	1	
$X_{E_r^\lambda}(So)$	0.9813	-0.9813	-1	1

TABLE 9. *pcc*-values of sixteen pairs of four measure sets on Wd.

pcc	$X_{G^\lambda}(Wd)$	$X_{E^\lambda}(Wd)$	$X_{H^\lambda}(Wd)$	$X_{E_r^\lambda}(Wd)$
$X_{G^\lambda}(Wd)$	1			
$X_{E^\lambda}(Wd)$	-1	1		
$X_{H^\lambda}(Wd)$	-0.9508	0.9508	1	
$X_{E_r^\lambda}(Wd)$	0.9508	-0.9508	-1	1

TABLE 10. The correlation between two measures on Ir.

	G^λ	E^λ	H^λ	E_r^λ
X_{G^λ}	CPC			
X_{E^λ}	CNC	CPC		
X_{H^λ}	HNC	HPC	CPC	
$X_{E_r^\lambda}$	HPC	HNC	CNC	CPC

TABLE 11. The correlation between two measures on Ec.

	G^λ	E^λ	H^λ	E_r^λ
X_{G^λ}	CPC			
X_{E^λ}	CNC	CPC		
X_{H^λ}	HNC	HPC	CPC	
$X_{E_r^\lambda}$	HPC	HNC	CNC	CPC

TABLE 12. The correlation between two measures on Pa.

	G^λ	E^λ	H^λ	E_r^λ
X_{G^λ}	CPC			
X_{E^λ}	CNC	CPC		
X_{H^λ}	HNC	HPC	CPC	
$X_{E_r^\lambda}$	HPC	HNC	CNC	CPC

TABLE 13. The correlation between two measures on Se.

	G^λ	E^λ	H^λ	E_r^λ
X_{G^λ}	CPC			
X_{E^λ}	CNC	CPC		
X_{H^λ}	HNC	HPC	CPC	
$X_{E_r^\lambda}$	HPC	HNC	CNC	CPC

TABLE 14. The correlation between two measures on So.

	G^λ	E^λ	H^λ	E_r^λ
X_{G^λ}	CPC			
X_{E^λ}	CNC	CPC		
X_{H^λ}	HNC	HPC	CPC	
$X_{E_r^\lambda}$	HPC	HNC	CNC	CPC

D. CORRELATION ANALYSIS

Define

$$pcc(S, T) = \frac{\sum_{i=1}^{n}(s_i - \bar{s})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^{n}(s_i - \bar{s})^2}\sqrt{\sum_{i=1}^{n}(t_i - \bar{t})^2}} \quad \text{(IV.2)}$$

where $S = \{s_1, \cdots, s_n\}$, $T = \{t_1, \cdots, t_n\}$, $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$, $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$.

$pcc(S, T)$ is called Pearson correlation coefficient between S and T. It reflects the degree of linear correlation between S and T.

The correlation between S and T can be obtained according to TABLE 3.

Continue the above experiment, *pcc*-values between two measurement sets are shown in TABLEs 4-9.

From TABLEs 4-9, the correlation degrees between the four measures on the six data sets are obtained (see TABLEs 10-15). Obviously, the six tables are the same. The correlation between the four measures on the six data sets is consistent.

E. FRIEDMAN TEST AND NEMENYI TEST

To further assess the performance of the proposed four measurement uncertainties, Friedman test and Nemenyi test are presented in this part.

TABLE 15. The correlation between two measures on Wd.

	G^λ	E^λ	H^λ	E_r^λ
X_{G^λ}	CPC			
X_{E^λ}	CNC	CPC		
X_{H^λ}	HNC	HPC	CPC	
$X_{E_r^\lambda}$	HPC	HNC	CNC	CPC

Friedman test is a statistical test on the basis of ranking method. It is defined as

$$\chi_F^2 = \frac{12N}{k(k+1)} \sum_{i=1}^{k} r_i^2 - 3N(k+1),$$

where N and k are the number of data sets and algorithms, respectively, r_i expresses the average ranking of the i-th algorithm. Nevertheless, the Friedman test is too conservation that is substituted by the following statistic

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2}.$$

The statistic F_F follows a Fisher distribution with $k-1$ and $(k-1)(N-1)$ degrees of freedom.

Nemenyi test is able to further explore which algorithm is better in the statistical term. And the critical distance CD_α is defined as

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \qquad \text{(IV.3)}$$

where q_α is the critical tabulated value and α is the significance level of the test. If the average level of distance exceeds CD_α, then the performance of two algorithms will be significantly different.

Continued from Subsection 4.2, we have

$G^\lambda(Ir) = 0.5058, E^\lambda(Ir) = 0.0238, H^\lambda(Ir) = 0.1808,$

$E_r^\lambda(Ir) = 0.3584, G^\lambda(Ec) = 0.6514, E^\lambda(Ec) = 0.1559,$

$H^\lambda(Ec) = 0.6409, E_r^\lambda(Ec) = 0.4145, G^\lambda(Pa) = 0.5869,$

$E^\lambda(Pa) = 0.0290, H^\lambda(Pa) = 0.2171, E_r^\lambda(Pa) = 0.3721,$

$G^\lambda(Se) = 0.3935, E^\lambda(Se) = 0.0100, H^\lambda(Se) = 0.1184,$

$E_r^\lambda(Se) = 0.2946, G^\lambda(So) = 0.3994, E^\lambda(So) = 0.0063,$

$H^\lambda(So) = 0.1030, E_r^\lambda(So) = 0.4119, G^\lambda(Wd) = 0.5484,$

$E^\lambda(Wd) = 0.0148, H^\lambda(Wd) = 0.1921, E_r^\lambda(Wd) = 0.3111,$

Four UMs can be seen as four algorithms. We demonstrate the statistical significance by using Friedman test and Nemenyi test.

We list the ranking of CV-values of the four measure sets on six datasets in TABLE 16.

Friedman test is used on four measurements under ten data sets. Thus, F_F follows the distribution with 3 and 15 degrees of freedom. It is noted that the critical value of Fisher distribution $F_{0.05}(3, 15)$ is 3.2874. Then $F_F = 35.9091$. Obviously, the value F_F is much bigger than 3.2874.

TABLE 16. The ranking of CV-values of the four measure sets on six data sets.

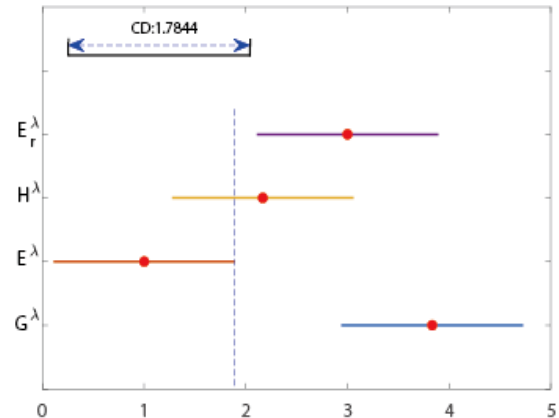| Date sets	G^λ	E^λ	H^λ	E_r^λ
Iris	4	1	2	3
Ecoli	4	1	3	2
Parkinsons	4	1	2	3
Seeds	4	1	2	3
Sona	3	1	2	4
Wdbc	4	1	2	3
Average	3.83	1.00	2.17	3.00

FIGURE 4. Nemenyi test.

This means that at the significant level $\alpha = 0.05$, the performance of the four measurements are different in the statistical significance.

The following results can be obtained from FIGURE 4,

1) Compare the performance of four measurements from statistics, E^λ is better than H^λ; In the same way, H^λ is better than E_r^λ and E_r^λ is better than G^λ.

2) There is no significant difference between the performance of E^λ and H^λ, H^λ and E_r^λ, E_r^λ and G^λ, respectively. But E^λ is significant difference from G^λ and E_r^λ.

F. PAIRED t-Test

The paired t-test may be regarded as an extension of the single-object t-test, except that Observations change from a group of independent objects of normal distribution to a pair of paired objects. For two paired objects $O_1 = \{o_{11}, o_{12}, \ldots, o_{1n}\}$ and $O_2 = \{o_{21}, o_{22}, \ldots, o_{2n}\}$, the differential between the paired observations is defined as $d_i = \{o_{1i} - o_{2i}\}$. If the differentials between the two pairs is independent of each other, and derived from normal distribution, then statistics below is able to be employed to determine whether the expectation of d_i is 0:

$$T = \frac{\bar{d}}{S_d/\sqrt{n}} \qquad \text{(IV.4)}$$

where $\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$, $S_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2}$, n is the number of paired objects. Under the null hypothesis, the statistic T follows t-distribution with degrees of freedom

$df = n - 1$. For a given significance level α, one can obtain rejection domain in the following:

$$W_t = \{|T| \geq t_{1-\frac{\alpha}{2}(n-1)}\} \qquad (IV.5)$$

The value of T computed by objects falls into the rejection domain that means two objects are significantly different. Since the smaller the CV-values, the better the measurement, we conduct paired t-tests on the CV-values. We treat the CV-values of G^λ and E^λ as a pair of paired objects. Likewise, we treat the CV-values of E^λ and H^λ, E^λ and E_r^λ, H^λ and E_r^λ, H^λ and G^λ, E_r^λ and G^λ as five pairs of paired objects, respectively. The tests are under the assumption that each pair objects come from normally distributed populations with unknown but equal variances. Pick $\alpha = 0.05$, the test results are shown in TABLE 17.

If $\bar{d} < 0$ and $p - value < 0.05$, it means the CV-values of the first measurement is significantly smaller than that of the second measurement. Thus, the performance of the first measurement is significantly better than that of the second. From TABLE 17, we obtain the following results:

a) The performance of E^λ is significantly better than that of G^λ, H^λ and E_r^λ;

b) The performance of G^λ is the worst of the four measures.

G. THE INFLUENCE OF PARAMETER λ ON THE FOUR MEASURES

The parameter λ is utilized to adjust the fuzzy relation; it has a significant influence on UM. We select an attribute subset $P = \{a_1, a_2, a_3\}$ for Ecoli data set and Parkinsons data set by adjusting the value of the parameter to vary from 0.1 to 0.9 with a step of 0.05. Then, we calculate the values of G^λ, E^λ, H^λ and E_r^λ. The results are shown in FIGUREs 5a-5d and FIGUREs 6a-6d.

From FIGUREs 5a-5d and FIGUREs 6a-6d, we obtain the following conclusions.

G^λ and E_r^λ are both monotonically decreasing as the parameter λ increases. By contraries, E^λ and H^λ are both monotonically increasing with the parameter λ growth. But they are not strictly monotonous.

V. AN APPLICATION IN ATTRIBUTE REDUCTION FOR AN IVIS

This part gives an application of the proposed measures in attribute reduction for an IVIS.

Definition 27: For an IVIS (O, AT), given $A \subseteq AT$. Then A is referred to as a λ-consistent subsets of AT, if $R_{AT}^\lambda = R_A^\lambda$.

The family of all λ-consistent subsets of AT is denoted by $co^\lambda(AT)$.

Definition 28: For an IVIS (O, AT), given $A \subseteq AT$. Then A is referred to as a λ-reduct of AT, if $A \in co^\lambda(AT)$ and $\forall\, a \in A$, $A - \{a\} \notin co^\lambda(AT)$.

The family of all λ-reducts of AT is denoted by $red^\lambda(AT)$.

Theorem 29: For an IVIS (O, AT), given $A \subseteq AT$. Then the following conditions are equivalent:
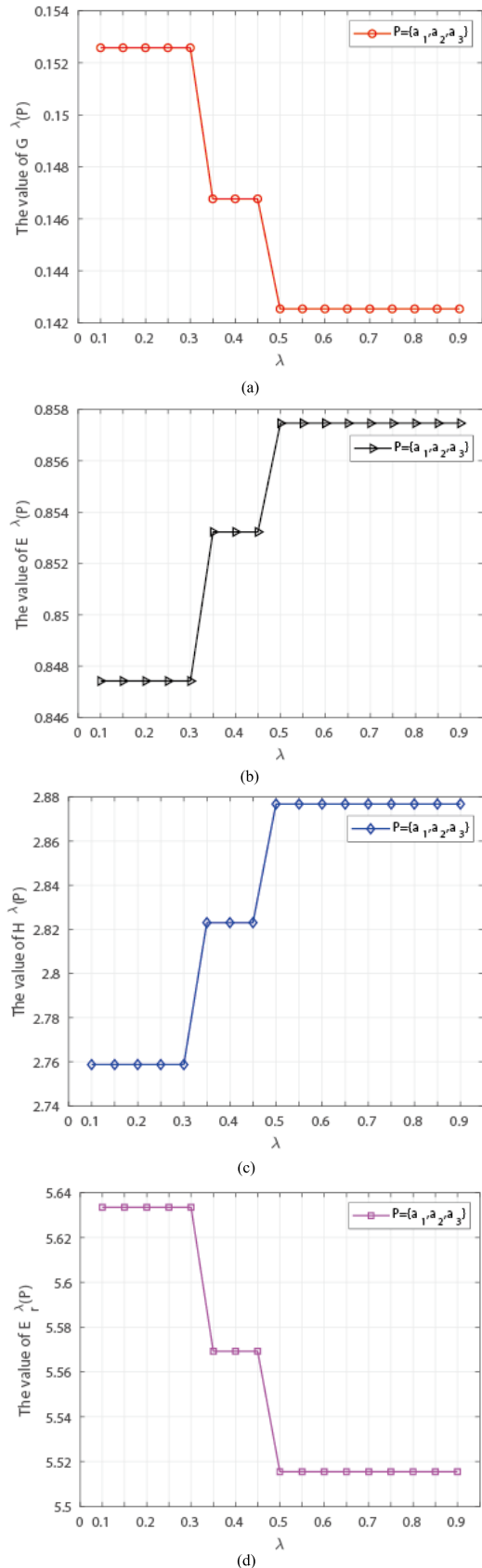
(1) $A \in co^\lambda(AT)$;

FIGURE 5. Paired t-test.

TABLE 17. Test results of paired *t*-test.

Pair	Measurement	$\bar{d}$	S_d	$t_{0.025}(5)$	$|T|$	df	$p-value$
1	E^λ vs. G^λ	−0.4743	0.0294	2.571	16.15	5	0.000
2	E^λ vs. H^λ	−0.2021	0.0585	2.571	3.453	5	0.018
3	E^λ vs. E_r^λ	−0.3205	0.0214	2.571	15.01	5	0.000
4	H^λ vs. E_r^λ	−0.1184	0.0738	2.571	1.604	5	0.170
5	H^λ vs. G^λ	−0.2722	0.0543	2.571	5.013	5	0.004
6	E_r^λ vs. G^λ	−0.1538	0.0401	2.571	3.832	5	0.012

Algorithm 1 Attribute Reduction Algorithm on the Basis of Fuzzy Information Granulation in an IVIS (FIEr-IVIS)

Input: An IVIS (O, AT).
Output: One reduct A.
Initialization: $A \leftarrow AT, flag = 1$.
Calculate $E_r^\lambda(AT)$.
while *flag* **do**
  for *each attribute* $a \in AT$ **do**
    if $E_r^\lambda(A - \{a\}) = E_r^\lambda(AT)$ **then**
      $A \leftarrow A - \{a\}$
    else
      $flag = 0$
      break
    end
  end
end
return A.

Algorithm 2 Attribute Reduction Algorithm Based on Fuzzy Information Entropy in an IVIS (FIE-IVIS)

Input: An IVIS (O, AT).
Output: One reduct A.
Initialization: $A \leftarrow \emptyset, flag = 1$.
Calculate $E^\lambda(AT)$.
while *flag* **do**
  for *each attribute* $a \in AT$ **do**
    if $E^\lambda(A \cup \{a\}) = E^\lambda(AT)$ **then**
      $A \leftarrow A \cup \{a\}$
    else
      $flag = 0$
      break
    end
  end
end
return A.

---

(2) $G^\lambda(A) = G^\lambda(AT)$;
(3) $H^\lambda(A) = H^\lambda(AT)$;
(4) $E_r^\lambda(A) = E_r^\lambda(AT)$;
(5) $E^\lambda(A) = E^\lambda(AT)$.

Proof: (1) $\Rightarrow$ (2). Clearly.
(2) $\Rightarrow$ (1). Suppose $G^\lambda(A) = G^\lambda(AT)$. Then

$$\frac{1}{n^2}\sum_{i=1}^{n}|R_A^\lambda(o_i)| = \frac{1}{n^2}\sum_{i=1}^{n}|R_{AT}^\lambda(o_i)|.$$

TABLE 18. Four data sets from UCI.

Date sets	Samples	Attributes	Classes
Wine	178	13	3
Leaf	340	15	36
Algerian Forest Fires	244	11	2
Climate	540	18	2

TABLE 19. Reduction results of four data sets from UCI.

Data sets	Reduction results
Wine	{1, 2, 4, 7, 8, 10, 11, 12, 13}
Leaf	{1,3,5,6,8,9,10,11,12,13,14,15}
Algerian Forest Fires	{1,2,3,4,5,7,8,10}
Climate	{1,2,3,4,6,8,9,10,11,13,14,16,18}

So

$$\sum_{i=1}^{n}(|R_{AT}^\lambda(o_i)| - |R_A^\lambda(o_i)|) = 0.$$

Note that $R_A^\lambda \subseteq R_{AT}^\lambda$. Then $\forall\ i,\ R_A^\lambda(o_i) \subseteq R_{AT}^\lambda(o_i)$. This implies that

$$\forall\ i,\ \ |R_{AT}^\lambda(o_i)| - |R_A^\lambda(o_i)| \geq 0.$$

So $\forall\ i,\ |R_{AT}^\lambda(o_i)| - |R_A^\lambda(o_i)| = 0$. It follows that $\forall\ i$, $R_{AT}^\lambda(o_i) = R_A^\lambda(o_i)$.
Thus $R_{AT}^\lambda = R_A^\lambda$. Hence

$$A \in co^\lambda(AT).$$

(2) $\Leftrightarrow$ (5). It can be proved by Theorem 25.
(1) $\Rightarrow$ (3). Clearly.
(3) $\Rightarrow$ (1). Suppose $H^\lambda(A) = H^\lambda(AT)$. Then
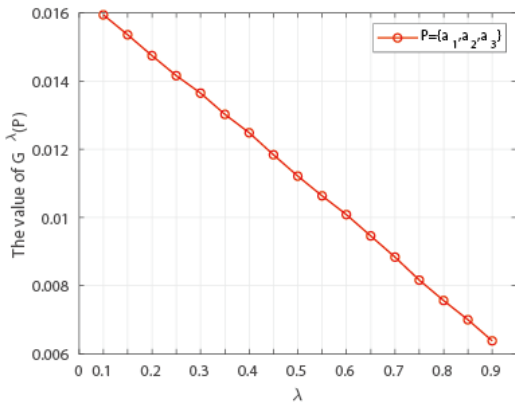
$$-\sum_{i=1}^{n}\frac{1}{n}\log_2\frac{|R_A^\lambda(o_i)|}{n} = -\sum_{i=1}^{n}\frac{1}{n}\log_2\frac{|R_{AT}^\lambda(o_i)|}{n}.$$
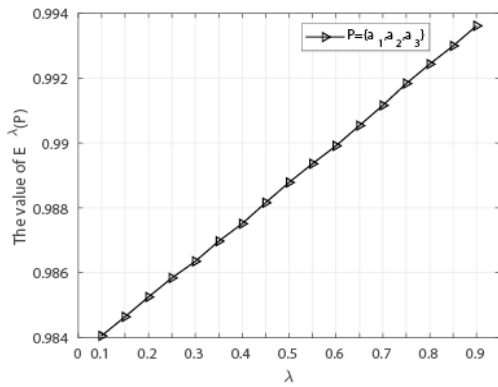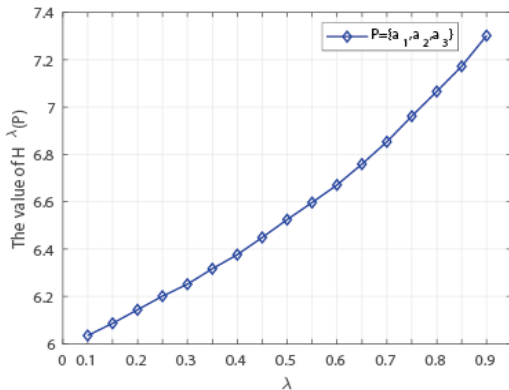
So

$$\sum_{i=1}^{n}\log_2\frac{|R_{AT}^\lambda(o_i)|}{|R_A^\lambda(o_i)|} = 0.$$

Note that $R_A^\lambda \subseteq R_{AT}^\lambda$. Then $\forall\ i,\ R_A^\lambda(o_i) \subseteq R_{AT}^\lambda(o_i)$. This implies that

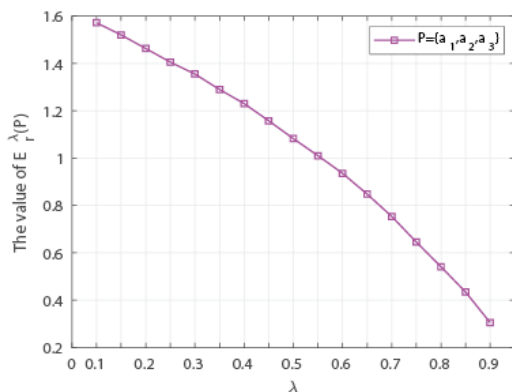$$\forall\ i,\ \ \log_2\frac{|R_{AT}^\lambda(o_i)|}{|R_A^\lambda(o_i)|} \geq 0.$$

(a)

(b)
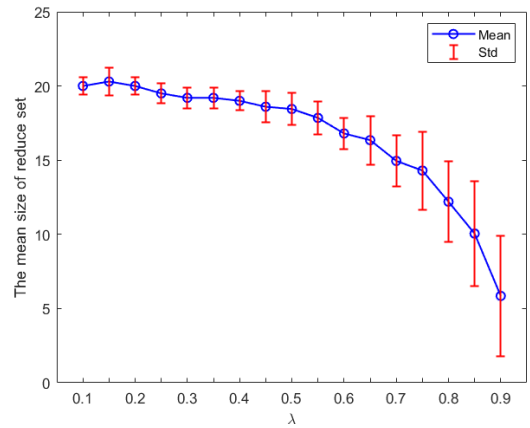

(c)

(d)

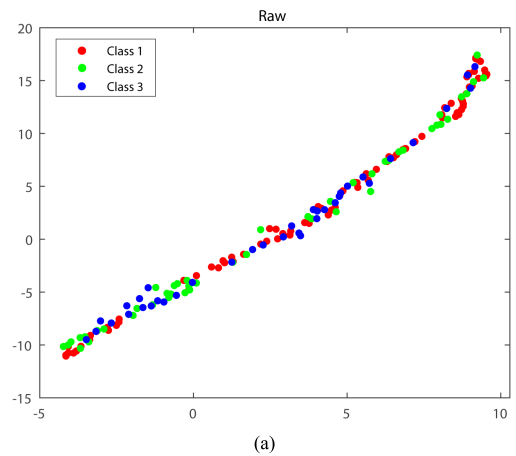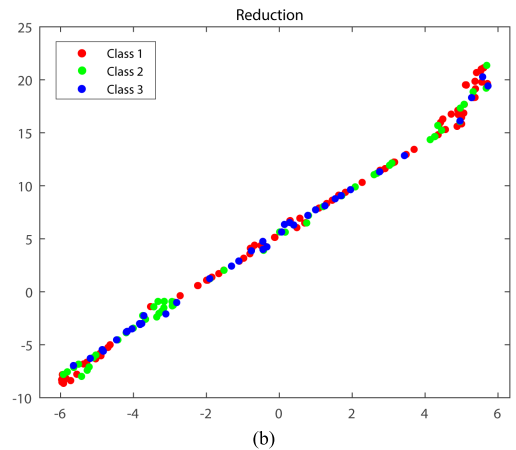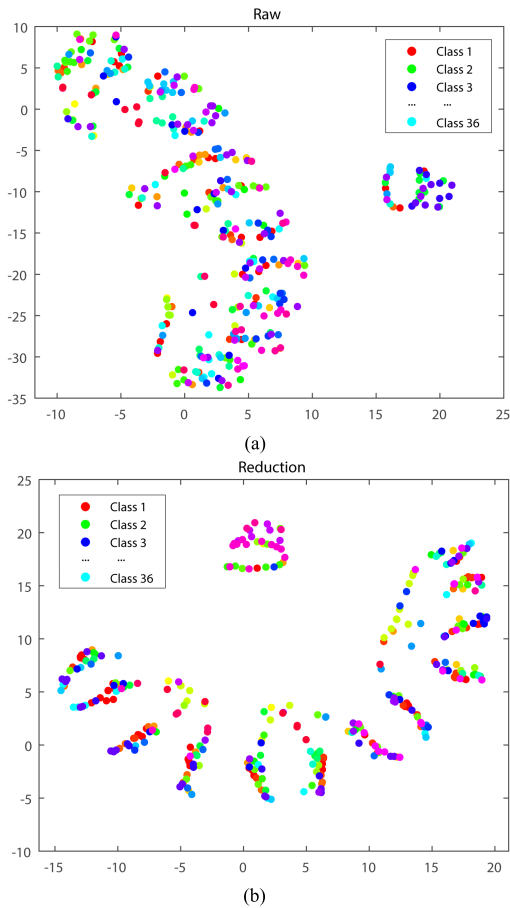**FIGURE 6.** Parameter test.

FIGURE 7. The influence of λ on Algorithm 1.

(a)

(b)

FIGURE 8. (a) The distribution before reduction on Wine. (b) The distribution after reduction on Wine.

So $\forall\ i,\ \log_2 \dfrac{|R_{AT}^\lambda(o_i)|}{|R_A^\lambda(o_i)|} = 0$. It follows that $\forall\ i,\ R_{AT}^\lambda(o_i) = R_A^\lambda(o_i)$.

Thus $R_{AT}^\lambda = R_A^\lambda$. Hence

$$A \in co^\lambda(AT).$$

$(4) \Leftrightarrow (3)$. It follows from Theorem 20. ∎

FIGURE 9. (a) The distribution before reduction on Leaf. (b) The distribution after reduction on Leaf.
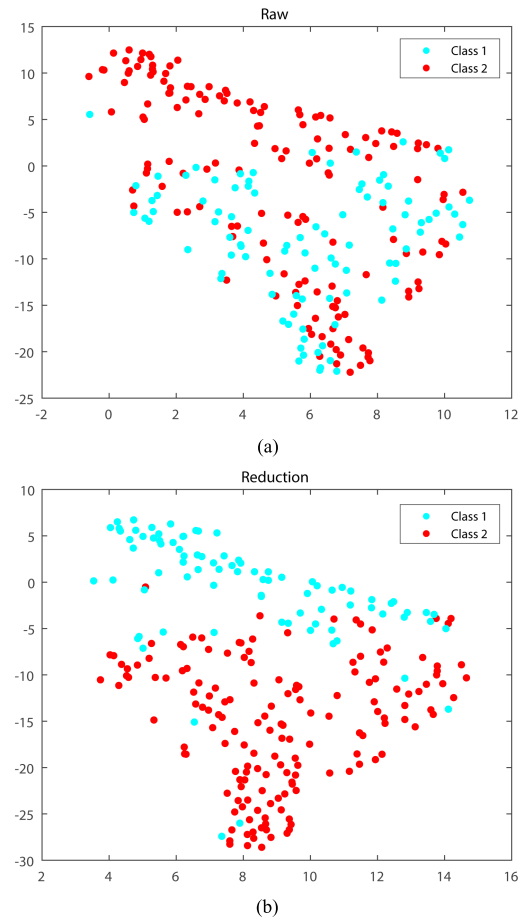

FIGURE 10. (a) The distribution before reduction on Algerian Forest Fires. (b) The distribution after reduction on Algerian Forest Fires.

Corollary 30: For an IVIS (O, AT), given $A \subseteq AT$. Then the following conditions are equivalent:

(1) $A \in red^\lambda(AT)$;
(2) $G^\lambda(A) = G^\lambda(AT)$ and $\forall a \in A, G^\lambda(A - \{a\}) \neq G^\lambda(AT)$;
(3) $H^\lambda(A) = H^\lambda(AT)$ and $\forall a \in A, H^\lambda(A - \{a\}) \neq H^\lambda(AT)$;
(4) $E_r^\lambda(A) = E_r^\lambda(AT)$ and $\forall a \in A, E_r^\lambda(A - \{a\}) \neq E_r^\lambda(AT)$;
(5) $E^\lambda(A) = E^\lambda(AT)$ and $\forall a \in A, E^\lambda(A - \{a\}) \neq E^\lambda(AT)$.

Proof: It is easy to be proved from Theorem 29. ∎

By Theorems 25 and 20, we have

$$G^\lambda(A) + E^\lambda(A) = 1, \quad E_r^\lambda(A) + H^\lambda(A) = log_2 \, n,$$

where (O, AT) is an IVIS and $A \subseteq AT$. Then, In the following, we only study the reduction algorithms on the basis of E_r^λ and E^λ and their overall time complexity are both $O(|A|^2)$.

A. CLUSTER ANALYSIS

In this subsection, clustering on the reduced IVIS is considered. We pick four data sets from UCI to do experiment (see TABLE 18). All real valued data in the these datasets must be converted into interval-valued data. The specific conversion method can be seen in Subsection 4.1. Each of these four data sets can be considered as an IVIS.

Below, we only use Algorithm 1 to get the reduction of each IVIS.

The parameter λ will have a certain impact on Algorithm 1. We first discuss the value of parameter λ and select Parkinsons data set by adjusting the value of the parameter λ to vary from 0.1 to 0.9 with a step of 0.05. After setting λ, Algorithm 1 is run for 20 times, and the mean and variance of the reduced subset size are counted. The results are shown in FIGURE 7.

From FIGURE 7, it can be seen that the reduced subset size is decreasing with the increase of λ, but it is not strictly decreasing, and its variance is also relatively large. To balance the reduction rate and the stability, it is recommended that λ is between 0.2 and 0.6 in practical application. In the following experiments of this paper, λ is set at 0.4.

These reduction results are shown in TABLE 19. From TABLE 19 show that Algorithm 1 can effectively reduce the dimension of data.

In order to verify the performance of the algorithm, Based on the work of Rouseeuw et al [25], we use k-medoids clustering algorithm to cluster the data sets before and after reduction.
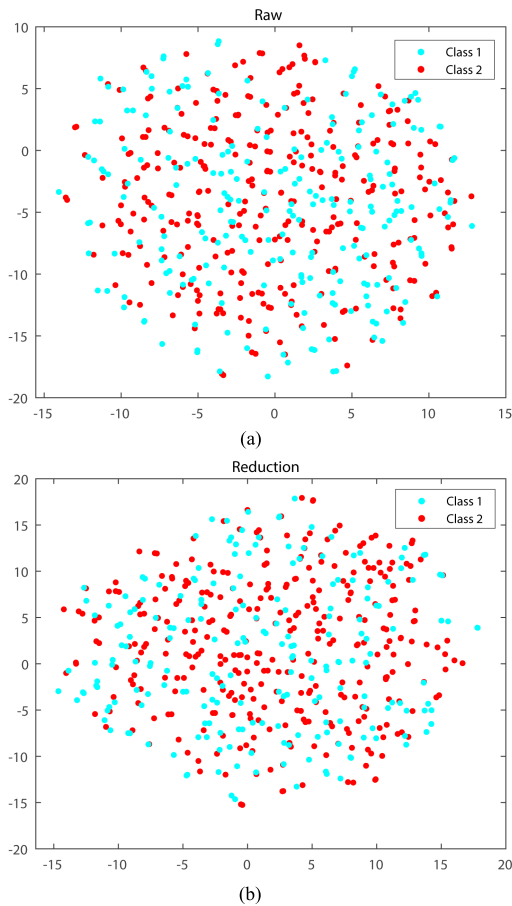
FIGURE 11. (a) The distribution before reduction on Climate. (b) The distribution after reduction on Climate.
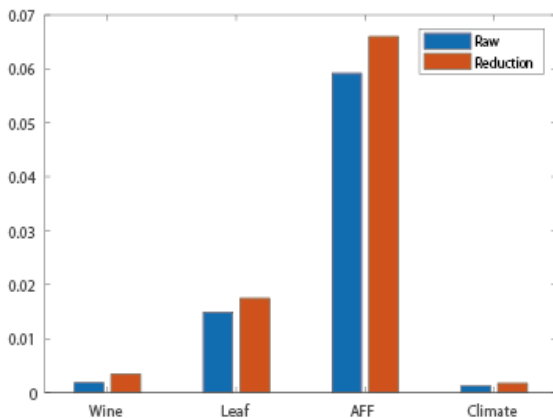

FIGURE 12. Silhouette coefficient.

The following distance between objects is to deal with interval-valued data.

Definition 31 [8]: For an IVIS (O, AT) with $AT = \{a_1, a_2, \cdots, a_m\}$, given $o, o' \in O$. Then the distance between o and o' is defined by

$$Dis(o, o') = \sqrt{\sum_{k=1}^{m}(|p(a_k(o), a_k(o')) - p(a_k(o'), a_k(o))|)^2}$$

(V.1)

It is worth noting that the number of clusters is set to the number of classes of the data set in all experiments. In order to intuitively display the clustering results, we take the midpoint of each interval to get the real value, and then reduce the dimension with PCA. The results are shown in FIGUREs 8a,8b,9a,9b,10a,10b,11a and 11b. The silhouette coefficient values of four data sets before and after reduction are shown in FIGURE 12. From FIGUREs 8a,8b,9a,9b,10a,10b,11a and 11b, it can be seen that the reduced data on these four data sets is more concentrated within the class and more dispersed among the classes. They indicate that the quality of reduced data is better. Figure 12 shows that the value of silhouette coefficient of the reduced data set is larger, indicating that the reduced data set has a better clustering effect.

VI. CONCLUSION

Considering that "The similarity between information values that is fed back to the attribute set", a new fuzzy symmetry relation on the object set of an IVIS has been constructed by introducing a variable parameter to control the similarity between information values. The advantage of this fuzzy symmetry relation is to facilitate fuzzy computation and then fuzzy Information granules have been obtained. Four UMs have been investigated by means of fuzzy information granules. The effectiveness of the investigated UMs has been proved by statistical analysis. The purpose of statistical analysis is to select two better measures to construct attribute reduction algorithm. Two attribute reduction algorithms on account of the selected UMs have been proposed. The experimental results of cluster analysis demonstrated the effectiveness of the proposed algorithm according to the evaluation indicator of clustering performance. In the future, we will apply the proposed measures for attribute reduction on large-scale gene data, and considers the possible application of our method in health-care waste treatment technologies assessment problems.

REFERENCES

[1] M. Akram, G. Ali, and J. C. R. Alcantud, "Parameter reduction analysis under interval-valued m-polar fuzzy soft information," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 5541–5582, Oct. 2021.

[2] G. Ali, M. Akram, and J. C. R. Alcantud, "Attributes reductions of bipolar fuzzy relation decision systems," *Neural Comput. Appl.*, vol. 32, no. 14, pp. 10051–10071, Jul. 2020.

[3] T. Beaubouef and F. E. Petry, "Fuzzy rough set techniques for uncertainty processing in a relational database," *Int. J. Intell. Syst.*, vol. 15, no. 5, pp. 389–424, May 2000.

[4] L. Chen, D. Chen, and H. Wang, "Fuzzy kernel alignment with application to attribute reduction of heterogeneous data," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 7, pp. 1469–1478, Jul. 2019.

[5] C. Cornelis, R. Jensen, G. Hurtado, and D. Slezak, "Attribute selection with fuzzy decision reducts," *Inf. Sci.*, vol. 180, no. 2, pp. 209–224, Jan. 2010.

[6] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 385–389, Apr. 2012.

[7] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, pp. 191–208, Jun. 1990.

[8] J. Dai, W. Wang, and J.-S. Mi, "Uncertainty measurement for interval-valued information systems," *Inf. Sci.*, vol. 251, pp. 63–78, Dec. 2013.

[9] J. Dai, W. Wang, Q. Xu, and H. Tian, "Uncertainty measurement for interval-valued decision systems based on extended conditional entropy," *Knowl.-Based Syst.*, vol. 27, pp. 443–450, Mar. 2012.

[10] J. Dai, B. Wei, X. Zhang, and Q. Zhang, "Uncertainty measurement for incomplete interval-valued information systems based on α-weak similarity," *Knowl.-Based Syst.*, vol. 136, pp. 159–171, Nov. 2017.

[11] I. Düntsch and G. Gediga, "Uncertainty measures of rough set prediction," *Artif. Intell.*, vol. 106, no. 1, pp. 109–137, Nov. 1998.

[12] A. Delgado and I. Romero, "Environmental conflict analysis using an integrated grey clustering and entropy-weight method: A case study of a mining project in Peru," *Environ. Model. Softw.*, vol. 77, pp. 108–121, Mar. 2016.

[13] G. Facchinetti, R. G. Ricci, and S. Muzzioli, "Note on ranking fuzzy triangular numbers," *Int. J. Intell. Syst.*, vol. 13, no. 7, pp. 613–622, Jul. 1998.

[14] Z. Huang and J. Li, "Discernibility measures for fuzzy β covering and their application," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9722–9735, Sep. 2022.

[15] G. Liu, Y. Feng, and J. Yang, "A common attribute reduction form for information systems," *Knowl.-Based Syst.*, vol. 193, Apr. 2020, Art. no. 105466.

[16] F. Li, C. Jin, and J. Yang, "Roughness measure based on description ability for attribute reduction in information system," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 5, pp. 925–934, May 2019.

[17] Z. Li, X. Liu, J. Dai, J. Chen, and H. Fujita, "Measures of uncertainty based on Gaussian kernel for a fully fuzzy information system," *Knowl.-Based Syst.*, vol. 196, May 2020, Art. no. 105791.

[18] Z. Li, L. Qu, G. Zhang, and N. Xie, "Attribute selection for heterogeneous data based on information entropy," *Int. J. Gen. Syst.*, vol. 50, no. 5, pp. 548–566, Jul. 2021.

[19] J. Liang, Z. Shi, D. Li, and M. J. Wierman, "Information entropy, rough entropy and knowledge granulation in incomplete information systems," *Int. J. Gen. Syst.*, vol. 35, no. 6, pp. 641–654, Dec. 2006.

[20] Z. Li, P. Zhang, X. Ge, N. Xie, G. Zhang, and C. Wen, "Uncertainty measurement for a fuzzy relation information system," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 12, pp. 2338–2352, Dec. 2019.

[21] J. Navarrete, D. Viejo, and M. Cazorla, "Color smoothing for RGB-D data using entropy information," *Appl. Soft Comput.*, vol. 46, pp. 361–380, Sep. 2016.

[22] Y. Nakahara, M. Sasaki, and M. Gen, "On the linear programming problems with set coefficients," *Comput. Ind. Eng.*, vol. 23, pp. 301–304, Nov. 1992.

[23] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982.

[24] Y. Qian, J. Liang, W.-Z.-Z. Wu, and C. Dang, "Information granularity in fuzzy binary GrC model," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 2, pp. 253–264, Apr. 2011.

[25] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[26] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[27] H. Sakai, M. Nakata, and D. Slezak, "A prototype system for rule generation in Lipski's incomplete information databases," in *Proc. 13th Rough Sets, Fuzzy Sets, Data Mining Granular Comput.*, 2011, pp. 175–182.

[28] 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, Mar. 2020.

[29] C. Wang, M. Qi, M. Shao, Q. Hu, D. Chen, Y. Qian, and Y. Lin, "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.

[30] C. Wang, Y. Wang, M. Shao, Y. Qian, and D. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 818–830, May 2020.

[31] Y. Wang, X. Chen, and K. Dong, "Attribute reduction via local conditional entropy," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 12, pp. 3619–3634, Dec. 2019.

[32] C. Wang, Y. Huang, M. Shao, Q. Hu, and D. Chen, "Feature selection based on neighborhood self-information," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 4031–4042, Sep. 2020.

[33] W. Wu, "Generalized fuzzy rough sets," *Inf. Sci.*, vol. 151, pp. 263–282, May 2003.

[34] L. Wang and H. Wang, "An integrated qualitative group decision-making method for assessing health-care waste treatment technologies based on linguistic terms with weakened hedges," *Appl. Soft Comput.*, vol. 117, Mar. 2022, Art. no. 108435.

[35] N. Xie, Z. Li, W.-Z. Wu, and G. Zhang, "Fuzzy information granular structures: A further investigation," *Int. J. Approx. Reasoning*, vol. 114, pp. 127–150, Nov. 2019.

[36] Y. Y. Yao, "Information granulation and rough set approximation," *Int. J. Intell. Syst.*, vol. 16, no. 1, pp. 87–104, Jan. 2001.

[37] Z. Yuan, H. Chen, T. Li, Z. Yu, B. Sang, and C. Luo, "Unsupervised attribute reduction for mixed data based on fuzzy rough sets," *Inf. Sci.*, vol. 572, pp. 67–87, Sep. 2021.

[38] L. Yang, X. Zhang, W. Xu, and 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, Sep. 2019.

[39] C. V. Negoita, "Fuzzy sets," *Fuzzy Sets Syst.*, vol. 133, no. 2, p. 275, Jan. 2003.

[40] Y. Zhang, T. Li, C. Luo, J. Zhang, and H. Chen, "Incremental updating of rough approximations in interval-valued information systems under attribute generalization," *Inf. Sci.*, vol. 373, pp. 461–475, Dec. 2016.

[41] G. Zhang, Z. Li, W.-Z. Wu, X. Liu, and N. Xie, "Information structures and uncertainty measures in a fully fuzzy information system," *Int. J. Approx. Reasoning*, vol. 101, pp. 119–149, Oct. 2018.

[42] J. Zeng, Z. Li, P. Zhang, and P. Wang, "Information structures and uncertainty measures in a hybrid information system: Gaussian kernel method," *Int. J. Fuzzy Syst.*, vol. 22, no. 1, pp. 212–231, Feb. 2020.

• • •