Information Entropy-Based Attribute Reduction for Incomplete Set-Valued Data

YUANXIA ZHANG AND ZUOZAN CHEN
School of Computer Science and Engineering, Yulin Normal University, Yulin, Guangxi 537000, China
Corresponding author: Zuozan Chen (zuozanchen100@126.com)

ABSTRACT This paper investigates attribute reduction for incomplete set-valued data based on information entropy. The similarity degree between information values on a conditional attribute of an incomplete set-valued decision information system (ISVDIS) is first proposed. Then, the tolerance relation on the object set with respect to a conditional attribute subset in an ISVDIS is obtained. Next, $\lambda$-reduction in an ISVDIS is presented. What’s more, connections between the proposed attribute reduction and uncertainty measurement are exhibited. Furthermore, an attribute reduction algorithm based on $\lambda$-information entropy in an ISVDIS is provided. Finally, experiments to evaluate the performance of the proposed algorithm are carried out, and Friedman test and Nemenyi test in statistics are conducted. The experimental results indicate that the proposed algorithm is more effective for an ISVDIS than some existing algorithms.

INDEX TERMS Rough set theory, ISVDIS, similarity degree, $\lambda$-reduction, $\lambda$-information entropy, attribute reduction algorithm.

I. INTRODUCTION
A. RESEARCH BACKGROUND
Rough set theory (RST) [26], [27] is a significant approach for managing uncertainty. An information system (IS) based on RST may reveal large databases and knowledge discovery process mathematically. Plenty of applications of RST are related to an IS [1], [2], [7], [10], [16], [19], [23], [33], [35]. An incomplete information system (IIS) expresses an IS including missing values. A set-valued information system (ISVIS) indicates an IS whose information values are sets. An incomplete set-valued information system (ISVIS) means an IIS whose information values are sets. An ISVIS with decision attributes is said to be an incomplete set-valued decision information system (ISVDIS). These ISs have been investigated by a great deal of scholars. For example, Yao et al. [44] proposed a RST model for a set-valued information system (SVIS) with upper and lower approximations and introduced generalized decision logic; Based on the process of knowledge induction, Leung et al. [20] came up with a rough set decision rule selection method based on minimum attribute set in a SVIS; Qian and Liang [28], [29] presented a dominance relation for a SVIS and a set-valued ordered IS.

The associate editor coordinating the review of this manuscript and approving it for publication was Francisco J. García-Penalvo.

Couso and Dubois [4] examined statistical reasoning with a SVIS from ontic and epistemic views, respectively. Chen et al. [5] measured the uncertainty of an ISVIS and considered the optimal selection of subsystems by using Gaussian kernel.

Uncertainty measurement may provide new viewpoint for analyzing data. We refer to the articles about uncertainty measurement of other scholars. For instance, Dai and Tian [8] thought about entropy measure and granularity measure of an SVIS; Wang and Yue [41] discussed entropy measure and granularity measure of interval and set-valued IS; Duntsch and Gediga [6] applied Shannon’s entropy to the measurement of decision rules in RST; Liu and Zhong [22] used four different kinds of entropy to gauge uncertainty in a fuzzy relation IS.

Attribute reduction in an ISVDIS is to delete some irrelevant attributes while maintaining the classification ability of the ISVDIS. As one of the core contents of RST, attribute reduction has been widely concerned. For instance, Guan et al. [15] studied attribute reduction in an ISVDIS and proposed the decision rules; Song an Zhang [31] proposed attribute reduction in a set-valued decision IS; Liu and Zhong [22] presented attribute reduction in a SVIS based on a dominance relation; Chen and Qin [3] discussed attribute reduction in a SVIS based on a tolerance relation; Li et al. [34] put forward attribute reduction in a ISVDIS.
TABLE 1. The comparison of this paper with other literatures.

| Literature          | Similarity degree | Binary relation          | Attribute selection method | Research tools |
|---------------------|-------------------|--------------------------|----------------------------|----------------|
| Dai et al. [9], 2013 | $S_{o'o'} = \frac{|b(o' \cap b(o')|}{|b(o) \cup b(o')|}$ | Fuzzy tolerance relation | Discernibility matrix | Fuzzy rough set |
| Liu et al. [22], 2016 | $c_{ij} = \frac{|f_i(o_j)|}{|f_i(o)|}$ | Dominance relation       | Discernibility matrix      | RST            |
| Singh et al. [32], 2019 | $s(o, o') = \frac{2|b(o) \cap b(o')|}{|b(o)| + |b(o')|}$ | Fuzzy tolerance relation | Degree of dependency      | RST            |

Differences

This article

1. $s(e(o), e(o')) = \frac{1}{\frac{|V_e|}{2}}$
2. $o \neq o'$, $e(o) = *, e(o') = *$
3. $o \neq o'$, $e(o) \neq *, e(o') = *$
4. $o \neq o'$, $e(o) = *, e(o') \neq *$
5. $o \neq o'$, $e(o) \neq *, e(o') \neq *, e(o) = e(o')$

Tolerance relation

Conditional entropy

RST

Similarities

1. All these articles have defined the similarity degree between set-valued attributes.
2. All these articles have investigated attribute reduction approaches for a SVDIS.
3. All these articles have took advantage of RST for an ISVDIS.

FIGURE 1. The work process of the article.

B. COMPARISON AND CONTRIBUTION

In order to see the innovation and contribution of this paper more clearly, we do comparison and discussion between this paper and some related literatures. They are shown in TABLE 1.

C. ORGANIZATION AND STRUCTURE

The work process of the article is displayed in FIGURE 1 and the rest is shown as below. Section 2 retrospects some essential notions of binary relations and ISVDISs. Section 3 constructs the similarity degree and tolerance relation in an ISVDIS, and considers rough approximations based on this tolerance relation. Section 4 proposes entropy measurement for an ISVDIS. Section 5 presents an attribute reduction algorithm based on $\lambda$-information entropy in an ISVDIS. Section 6 gives an illustrative example. Section 7 evaluates the performance of the presented algorithm and does some statistical hypothesis experiments. Section 8 concludes this article.

II. PRELIMINARIES

In this section, we recall some basic notions about binary relations and ISVDISs.

Throughout this paper, $O$ denotes a finite set, $2^O$ means the family of all subsets of $O$ and $|X|$ expresses the cardinality of $X \in 2^O$. Put

$$O = \{o_1, o_2, \cdots, o_n\}, \delta = O \times O, \Delta = \{(o, o) : o \in O\}.$$

A. BINARY RELATIONS

Recall that $R$ is a binary relation on $O$ whenever $R \subseteq O \times O$.

If $(o, o') \in R$, then we denote it by $oR o'$.

Let $R$ be a binary relation on $O$. Then $R$ is said to be

1. Reflexive, if $\forall o \in O, oR o$;
2. Symmetric, if $oR o'$ implies $o' R o$;
3. Transitive, if $oR o'$, $o' R o''$ imply $oR o''$.

$R$ is said to be an equivalence relation on $O$, if $R$ is reflexive, symmetric and transitive. $R$ is said to be a tolerance relation on $O$, if $R$ is reflexive and symmetric. Moreover, $R$ is said to be a universal relation on $O$ if $R = \Delta$; $R$ is said to be an identity relation on $O$ if $R = \delta$.

B. AN ISVDIS

Definition 1 [27]: Let $O$ be a finite object set and $E$ a finite attribute set. Then $(O, E)$ is a called an information system (IS), if $\forall e \in E, e$ determines an information function $a : O \rightarrow V_e$, where $V_e = \{e(o) : o \in O\}$.

Let $(O, E)$ be an IS. If $\exists e \in E, * \in V_e$, here * means a unknown value, then $(O, E)$ is said to be an incomplete information system (IIS).
Let \((O, E)\) be an IIS. \(\forall e \in E\), denote

\[ V^e_e = V_e - \{e(o) : e(o) = \ast\}. \]

Then, \(V^e_e\) means the set of all non-missing information values of the attribute \(e\).

Let \((O, E)\) be an IS. If \(E = C \cup D\) where \(C\) is a set of conditional attributes and \(D\) is a set of decision attributes, then \((O, E)\) is said to be a decision information system (DIS).

**Definition 2** [43]: Suppose that \((O, E)\) is an IS. Then \((O, E)\) is referred to as a set-valued information system (SVIS), if \(\forall e \in E, \forall o \in O, (e(o)\) is a set.

**Definition 3** [43]: Given that \((O, C \cup \{d\})\) is a DIS. Then \((O, C \cup \{d\})\) is said to be an incomplete set-valued decision information system (ISVDIS), if \((O, C \cup \{d\})\) is both incomplete and set-valued.

If \(F \subseteq C\), then \((O, F \cup \{d\})\) is referred to as the subsystem of \((O, C \cup \{d\})\).

**Example 1**: TABLE 2 depicts an ISVDIS \((O, C \cup \{d\})\), where \(O = \{o_1, o_2, \ldots, o_{12}\}\) and \(E = \{e_1, e_2, \ldots, e_8\}\).

**Example 2 (Continued From Example 1)**:

\[
\begin{align*}
V^*_{e_1} &= \{\{1\}, \{2\}\}, & V^*_{e_2} &= \{\{1\}, \{2\}, \{3\}\}, \\
V^*_{e_3} &= \{\{1\}, \{2\}, \{3\}\}, & V^*_{e_4} &= \{\emptyset, \{2\}\}, \\
V^*_{e_5} &= \{\emptyset, \{2\}\}, & V^*_{e_6} &= \{\{1\}, \{3\}\}, \\
V^*_{e_7} &= \{\{1\}, \{3\}\}, & V^*_{e_8} &= \{\{1\}, \{2\}, \{3\}\}. \\
\end{align*}
\]

\(O/d = \{D_1, D_2, D_3\}\),

where \(D_1 = \{o_1, o_6, o_8, o_{10}\}\), \(D_2 = \{o_2, o_3, o_5, o_9, o_{11}\}\), \(D_3 = \{o_4, o_7, o_{12}\}\).

### III. THE TOLERANCE RELATIONS INDUCED BY AN ISVDIS AND ROUGH APPROXIMATIONS BASED ON THEM

In this part, we give the tolerance relations induced by an ISVDIS and define rough approximations based on them.

#### A. THE SIMILARITY DEGREE BETWEEN INFORMATION VALUES ON A CONDITIONAL ATTRIBUTE

**Definition 4**: Let \((O, C \cup \{d\})\) be an ISVDIS. Then \(\forall o, o' \in O, e \in C\), the similarity degree between \(e(o)\) and \(e(o')\) is defined as follows:

\[
s(e(o), e(o')) = \begin{cases} 
1, & o = o' \\
\frac{1}{|V^e_e|}, & o \neq o', e(o) = \ast, e(o') = \ast; \\
\frac{1}{|V^e_e|}, & o \neq o', e(o) \neq \ast, e(o') = \ast; \\
\frac{1}{|V^e_e|}, & o \neq o', e(o) = \ast, e(o') \neq \ast; \\
\frac{|e(o) \cap e(o')|}{|e(o) \cup e(o')|}, & o \neq o', e(o) \neq \ast, e(o') \neq \ast, e(o) = e(o'); \\
\frac{|e(o) \cap e(o')|}{|e(o) \cup e(o')|} & o \neq o', e(o) \neq \ast, e(o') \neq \ast, e(o) \neq e(o').
\end{cases}
\]

Put

\[ s^k_{ij} = s(e_k(o_i), e_k(o_j)). \]

**Example 3 (Continued From Example 1)**: \(\forall i, j, k, s^k_{ij}\) is obtained as follows (see TABLES 3-10).

#### B. TOLERANCE RELATIONS IN AN ISVDIS

**Definition 5**: Consider that \((O, C \cup \{d\})\) is an ISVDIS. Given \(\lambda \in [0, 1]\) and \(F \subseteq C\). Define

\[
R^F_F = \{(o, o') \in O \times O : \forall e \in F\}
\]

\[ s(e(o), e(o')) \geq \lambda, \]

\[
T_d = \{(o, o') \in O \times O : d(o) = d(o')\}.
\]

Obviously, \(R^F_F\) and \(T_d\) are tolerance and equivalence relations on \(O\), respectively.

For \(\forall o \in O\), define

\[
R^F_F(o) = \{o' \in O : (o, o') \in R^F_F\}
\]

\[
T_d(o) = \{o' \in O : d(o) = d(o')\}.
\]

Then \(R^F_F(o)\) and \(T_d(o)\) are said to be the \(\lambda\)-tolerance and equivalence classes of \(o\) under \(R^F_F\) and \(T_d\), respectively. Moreover, \(T_d(o)\) is said to be the decision class of \(o\) in \(C\), and \(T_d\) is said to be the decision in \(C\).

In this paper, denote

\[
O/d = \{T_d(o) : o \in O\} = \{D_1, D_2, \ldots, D_r\},
\]

where \(r = |V_d|\).

**Example 4 (Continued From Example 3)**: Pick \(\lambda = 0.4\). Then \(\forall i, \forall o \in O, \) the tolerance class \(R^F_F(o)\) of \(o\) is obtained (see TABLES 11-13).

An algorithm for computing the \(\lambda\)-tolerance class is designed as follows.

**Definition 6**: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(\lambda \in [0, 1]\) and \(F \subseteq C\). Define \(\delta_F^F : O \rightarrow 2^F\lambda\) as follows:

\[
\delta_F^F(o) = \{d(o') \in \delta_F^F(o)\}.
\]

Then \(\delta_F^F(o)\) is said to be \(\lambda\)-generalized decision of \(o\) in \(C\), and \(\delta_F^F\) is said to be \(\lambda\)-generalized decision in \(C\).

**Proposition 1**: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(\lambda \in [0, 1]\), \(F \subseteq C\) and \(o \in O\). Then

\[
R^F_F(o) \subseteq T_d(o) \iff |\delta_F^F(o)| = 1.
\]

**Proof**: \(\Rightarrow\). Let \(R^F_F \subseteq T_d(o)\). Suppose \(z \in \delta_F^F(o)\). Then there exists \(o' \in R^F_F\) such that \(z = d(o')\). \(o' \in R^F_F\) implies that \(o' \in T_d(o)\). So \(z = d(o') = d(o)\). Thus \(|\delta_F^F(o)| = 1\).

\(\Leftarrow\). Let \(|\delta_F^F(o)| = 1\). Suppose \(o' \in \delta_F^F(o)\). Then \(d(o') \in \delta_F(o)\). Since \(d(o) \in \delta_F^F(o)\) and \(|\delta_F^F(o)| = 1\), we have \(d(o) = d(o')\). Then \(o' \in T_d(o)\). Thus \(R^F_F(o) \subseteq T_d(o)\).

**Corollary 1**: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(\lambda \in [0, 1]\) and \(F \subseteq C\). Then

\[
R^F_F \subseteq T_d \iff \forall o \in O, \ |\delta_F^F(o)| = 1.
\]

**Proof**: It can be obtained by Proposition 1.
TABLE 2. An ISVDIS \((O, C \cup \{d\})\).

| \(e_1\) | \(e_2\) | \(e_3\) | \(e_4\) | \(e_5\) | \(e_6\) | \(e_7\) | \(e_8\) | \(d\) |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| \(o_1\) | \(\{1, 2\}\) | \(\{1, 2\}\) | \(\{1, 2, 3\}\) | \(\emptyset\) | \(\{2\}\) | \(\{2, 3\}\) | \(\{1\}\) | 1 |
| \(o_2\) | \(\{2\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | 2 |
| \(o_3\) | \(\{2\}\) | \(\{2, 3\}\) | \(\emptyset\) | \(\{1, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | \(\{2, 3\}\) | 2 |
| \(o_4\) | \(\{2\}\) | \(\{1, 2\}\) | \(\emptyset\) | \(\{2\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | 2 |
| \(o_5\) | \(\{1\}\) | \(\{1, 2\}\) | \(\emptyset\) | \(\{2\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | 2 |
| \(o_6\) | \(\{2\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | 2 |
| \(\rho_{10}\) | \(\{2\}\) | \(\{1, 2\}\) | \(\emptyset\) | \(\{2\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | 2 |
| \(o_{11}\) | \(\{2\}\) | \(\{1, 2\}\) | \(\emptyset\) | \(\{2\}\) | \(\{2, 3\}\) | \(\{1, 3\}\) | \(\{2, 3\}\) | 2 |
| \(\rho_{12}\) | \(\{1\}\) | \(\{1, 3\}\) | \(\{1, 2, 3\}\) | \(\{1\}\) | \(\{1, 3\}\) | \(\{1, 2, 3\}\) | \(\{1\}\) | 3 |

TABLE 3. \(s_{ij}^1\).

| \(s_{i,j}^1\) | \(o_1\) | \(o_2\) | \(o_3\) | \(o_4\) | \(o_5\) | \(o_6\) | \(o_{10}\) | \(o_{11}\) | \(o_{12}\) |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| \(o_1\) | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.25 | 0.5 | 0.5 |
| \(o_2\) | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| \(o_3\) | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| \(o_4\) | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| \(o_5\) | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| \(o_6\) | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| \(o_{10}\) | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| \(o_{11}\) | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| \(o_{12}\) | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

TABLE 4. \(s_{ij}^2\).

| \(s_{i,j}^2\) | \(o_1\) | \(o_2\) | \(o_3\) | \(o_4\) | \(o_5\) | \(o_6\) | \(o_{10}\) | \(o_{11}\) | \(o_{12}\) |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| \(o_1\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_2\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_3\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_4\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_5\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_{10}\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_{11}\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| \(o_{12}\) | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |

Definition 7: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(\lambda \in [0, 1]\). If \(\forall o \in O, |\partial_{\lambda}^C(o)| = 1\), then \((O, C \cup \{d\})\) is said to be \(\lambda\)-consistent; otherwise, \((O, C \cup \{d\})\) is said to be \(\lambda\)-inconsistent.

Theorem 1: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(\lambda \in [0, 1]\). Then \((O, C \cup \{d\})\) is \(\lambda\)-consistent \(\iff R_{\lambda}^C \subseteq T_d\).

C. ROUGH APPROXIMATIONS BASED ON THE TOLERANCE RELATIONS INDUCED BY AN ISVDIS

Definition 8: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Based on the approximation space \((O, R_{\lambda}^F)\), define \(R_{\lambda}^C, R_{\lambda}^F: 2^O \rightarrow 2^O\) as follows:

\[ R_{\lambda}^C(X) = \{o \in O : R_{\lambda}^C(o) \subseteq X\}, \]
\[ R_{\lambda}^F(X) = \{o \in O : R_{\lambda}^F(o) \cap X \neq \emptyset\}. \]
Theorem 2: Let \( (O, C \cup \{d\}) \) be an ISVDIS.

(1) \( R^\lambda_\nu(X) = R^\lambda_\nu(\emptyset) = \emptyset \); \( R^\lambda_\nu(O) = R^\lambda_\nu(O) = O \).

(2) \( \bar{R}^\lambda_\nu(X) \subseteq X \subseteq \bar{R}^\lambda_\nu(X) \).

Then \( R^\lambda_\nu(X) \) and \( \bar{R}^\lambda_\nu(X) \) are said to be \( \lambda \)-lower and \( \lambda \)-upper approximations of \( X \), respectively.

Theorem 2: Let \( (O, C \cup \{d\}) \) be an ISVDIS.

(1) \( R^\lambda_\nu(O) = \bar{R}^\lambda_\nu(O) = O \).

(2) \( R^\lambda_\nu(X) \subseteq X \subseteq \bar{R}^\lambda_\nu(X) \).

Proof: It can be obtained by Theorem 2.

Definition 10: Let \( (O, C \cup \{d\}) \) be an ISVDIS. Given \( F \subseteq C \) and \( \lambda \in [0, 1] \). Then \( \lambda \)-dependence degree of \( d \) over \( F \) is defined as

\[
\Gamma^\lambda_f(d) = \frac{\sum_{D \in O/d} |R^\lambda_\nu(F(D))|}{n}.
\]

Theorem 4: Let \( (O, C \cup \{d\}) \) be an ISVDIS. Denote \( O/d = \{D_1, D_2, \ldots, D_r\} \).

(1) \[ O/d = \{D_1, D_2, \ldots, D_r\} \]
The tolerance class of each object under $R_F^*(e = e_1, e_2, e_3, \lambda = 0.4)$. 

| $R_{O_1}^*(o)$ | $R_{O_2}^*(o)$ | $R_{O_3}^*(o)$ |
|---------------|---------------|---------------|
| $\{o_1\}$    | $\{o_2, o_3, o_{10}\}$ | $\{o_2, o_3, o_{10}\}$ |
| $\{o_2\}$    | $\{o_1, o_{10}\}$ | $\{o_1, o_{10}\}$ |
| $\{o_3\}$    | $\{o_1, o_{10}\}$ | $\{o_1, o_{10}\}$ |
| $\{o_4\}$    | $\{o_2, o_{10}\}$ | $\{o_2, o_{10}\}$ |
| $\{o_5, o_{10}\}$ | $\{o_2, o_{10}\}$ | $\{o_2, o_{10}\}$ |
| $\{o_6\}$    | $\{o_1, o_{10}\}$ | $\{o_1, o_{10}\}$ |
| $\{o_7\}$    | $\{o_1, o_{10}\}$ | $\{o_1, o_{10}\}$ |
| $\{o_8\}$    | $\{o_1, o_{10}\}$ | $\{o_1, o_{10}\}$ |
| $\{o_9\}$    | $\{o_1, o_{10}\}$ | $\{o_1, o_{10}\}$ |
| $\{o_{10}\}$ | $\{o_1, o_2\}$ | $\{o_1, o_2\}$ |
| $\{o_1, o_2\}$ | $\{o_2, o_3, o_{10}\}$ | $\{o_2, o_3, o_{10}\}$ |

Algorithm 1 Computing $R_F^*(o)$

Input: An ISVDIS $(O, C \cup \{d\})$, a threshold $\lambda \in [0, 1]$, $F \subseteq C$, $\lambda$, and $o \in O$.

Output: The tolerance class $R_F^*(o)$.

1. for $o' \in O$ do
   2. for $v = u$, do
      3. $s(e(o), e(o')) = 1$.
   4. end
   5. for $v \neq u$ do
      6. if $e(o) = *$ and $e(o') = *$, then
         7. $s(e(o), e(o')) = \frac{1}{|R_F^*(o)|}$.
      8. end
      9. if $e(o) = *$ and $e(o') \neq *$, then
         10. $s(e(o), e(o')) = \frac{1}{|R_F^*(o)|}$.
      11. end
      12. if $e(o) \neq *$ and $e(o') = *$, then
         13. $s(e(o), e(o')) = \frac{1}{|R_F^*(o)|}$.
      14. end
      15. if $e(o) \neq *$ and $e(o') \neq *$ and $e(o) = e(o')$, then
         16. $s(e(o), e(o')) = 1$.
      17. end
      18. if $e(o) \neq *$ and $e(o') \neq *$ and $e(o) \neq e(o')$, then
         19. $s(e(o), e(o')) = \forall(e(o), e(o'))$.
      20. end
   21. end
22. end
23. Put $R_F^*(o) = \{(o, o') \in O \times O : \forall e \in F, s(e(o), e(o')) \geq \lambda\}$.
24. Obtain $R_F^*(o)$.

Proof: (1) By Theorem 2(2), $\forall j, R_F^*(D_j) \subseteq D_j$.
Note that $\{D_1, D_2, \ldots, D_r\}$ is a partition of $O$. Then

$$|POS_F^*(D)| = \bigcup_{j=1}^{r} R_F^*(D_j) = \sum_{j=1}^{r} |R_F^*(D_j)|.$$ 

Thus

$$\Gamma_F^*(d) = \frac{\sum_{j=1}^{r} |R_F^*(D_j)|}{n}.$$ 

(2) It follows from (1) and Theorem 3(2).
(3) This holds by Theorem 3(4).
(4) It can be obtained by Theorem 3(5). □

IV. ENTROPY MEASUREMENT FOR AN ISVDIS

In this section, we propose entropy measurement for an ISVDIS.

Stipulate $0 \log_2 0 = 0$.

Definition 11: Suppose that $(O, C \cup \{d\})$ is an ISVDIS.
Given $F \subseteq C$ and $\lambda \in [0, 1]$. Then $\lambda$-information entropy of $F$ is defined as

$$H^\lambda(F) = -\sum_{i=1}^{n} \frac{|R_F^*(o_i)|}{n} \log_2 \frac{|R_F^*(o_i)|}{n}.$$ 

Proposition 2: Let $(O, C \cup \{d\})$ be an ISVDIS. Given $F \subseteq C$ and $\lambda \in [0, 1]$. Then

$$0 \leq H^\lambda(F) \leq \log_2 n.$$ 

Moreover, if $R_F^* = \Delta$, then $H^\lambda(F) = \log_2 n$; if $R_F^* = \delta$ is a universal relation on $O$, then $H^\lambda$ achieves the minimum value 0.

Proof: Since $\forall i, 1 \leq |R_F^*(o_i)| \leq n$, we have

$$\frac{1}{n} \leq \frac{|R_F^*(o_i)|}{n} \leq 1,$n \leq \log_2 \frac{|R_F^*(o_i)|}{n} \leq \log_2 n.$$ 

Then

$$0 \leq -\log_2 \frac{|R_F^*(o_i)|}{n} \log_2 \frac{|R_F^*(o_i)|}{n} \leq \log_2 n.$$
Definition 12: Let \( R_\lambda \) be an ISVDIS. Given \( F \subseteq C \) and \( \lambda \in [0, 1] \), then \( \lambda \)-conditional information entropy of \( F \) to \( d \) is defined as

\[
H^\lambda(d|F) = - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R^\lambda_\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R^\lambda_\lambda(o_i) \cap D_j|}{|R^\lambda_\lambda(o_i)|}. 
\]

Proposition 3: Assume that \((U, C \cup \{d\})\) is an ISVDIS.

1. If \( T \subseteq S \subseteq C \), then \( \forall \lambda \in [0, 1] \),

\[
H^\lambda(S|d) \leq H^\lambda(T|d). 
\]

2. If \( 0 < \lambda_1 \leq \lambda_2 \leq 1 \), then \( \forall F \subseteq C \),

\[
H^{\lambda_1}(F|d) \leq H^{\lambda_2}(F|d). 
\]

Proof: (1) Denote

\[
\begin{align*}
\tilde{t}_{ij}^{(1)} &= |R^\lambda_\lambda(o_i) \cap D_j|, \\
\tilde{t}_{ij}^{(2)} &= |R^\lambda_\lambda(U \cap D_j)|, \\
\tilde{s}_{ij}^{(1)} &= |R^\lambda_\lambda(o_i) \cap (U - D_j)|, \\
\tilde{s}_{ij}^{(2)} &= |R^\lambda_\lambda(U - D_j)|.
\end{align*}
\]

Then

\[
|R^\lambda_\lambda(o_i)| = \tilde{t}_{ij}^{(1)} + \tilde{t}_{ij}^{(2)}; \\
|R^\lambda_\lambda(o_i)| = \tilde{s}_{ij}^{(1)} + \tilde{s}_{ij}^{(2)}.
\]

Obviously, \( \forall i, R^\lambda_\lambda(o_i) \subseteq R^\lambda_\lambda(o_i) \). Then

\[
\begin{align*}
\forall i, j, 0 \leq \tilde{s}_{ij}^{(1)} \leq \tilde{t}_{ij}^{(1)}, & \quad 0 \leq \tilde{s}_{ij}^{(2)} \leq \tilde{t}_{ij}^{(2)}. \\
H^\lambda(T|d) &= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R^\lambda_\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R^\lambda_\lambda(o_i) \cap D_j|}{|R^\lambda_\lambda(o_i)|} \\
&= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{\tilde{t}_{ij}^{(1)} + \tilde{s}_{ij}^{(1)}}{n} \log_2 \frac{\tilde{t}_{ij}^{(1)} + \tilde{s}_{ij}^{(2)}}{\tilde{t}_{ij}^{(1)} + \tilde{t}_{ij}^{(2)}} \\
&\leq - \sum_{i=1}^{n} \sum_{j=1}^{r} f(\tilde{t}_{ij}^{(1)}, \tilde{t}_{ij}^{(2)}).
\end{align*}
\]

\[
\begin{align*}
H^\lambda(S|d) &= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R^\lambda_\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R^\lambda_\lambda(o_i) \cap D_j|}{|R^\lambda_\lambda(o_i)|} \\
&= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{\tilde{s}_{ij}^{(1)} + \tilde{s}_{ij}^{(2)}}{\tilde{s}_{ij}^{(1)} + \tilde{s}_{ij}^{(2)}} \\
&\leq - \sum_{i=1}^{n} \sum_{j=1}^{r} f(\tilde{s}_{ij}^{(1)}, \tilde{s}_{ij}^{(2)}).
\end{align*}
\]

Put \( f(x, y) = -x \log_2 \frac{1}{x+y}(x > 0, y \geq 0) \). Then \( f(x, y) \) increases with respect to \( x \) and increases with respect to \( y \), respectively.

Since \( \tilde{s}_{ij}^{(1)} \leq \tilde{t}_{ij}^{(1)}, \tilde{s}_{ij}^{(2)} \leq \tilde{t}_{ij}^{(2)} \), we have

\[
\tilde{t}_{ij}^{(1)}, \tilde{t}_{ij}^{(2)} \leq f(\tilde{t}_{ij}^{(1)}, \tilde{t}_{ij}^{(2)}) \leq f(\tilde{s}_{ij}^{(1)}, \tilde{s}_{ij}^{(2)}).
\]

Thus

\[
H^\lambda(S|d) \leq H^\lambda(T|d).
\]

(2) Denote

\[
\begin{align*}
p_{ij}^{(1)} &= |R^\lambda_\lambda(o_i) \cap D_j|, \\
p_{ij}^{(2)} &= |R^\lambda_\lambda(o_i) \cap (U - D_j)|, \\
q_{ij}^{(1)} &= |R^\lambda_\lambda(o_i) \cap D_j|, \\
q_{ij}^{(2)} &= |R^\lambda_\lambda(o_i) \cap (U - D_j)|.
\end{align*}
\]

Obviously, \( \forall i, R^\lambda_\lambda(o_i) \subseteq R^\lambda_\lambda(o_i) \).
Then
\[
\forall i, j, \ 0 \leq p_{ij}^{(1)} \leq q_{ij}^{(1)}, \ 0 \leq p_{ij}^{(2)} \leq q_{ij}^{(2)}.
\]
\[
p_{ij}^{(1)} + p_{ij}^{(2)} = \sum_{v \in D_j} G_F^1(o_i, v) + \sum_{v \in U \setminus D_j} G_F^2(o_i, v)
= \sum_{v \in U} G_F^2(o_i, v)
= \sum_{v \in U} R_F^2(o_i, v)
= |R_F^2(o_i)|.
\]
Similarly, it can be obtained that \( \forall i, q_{ij}^{(1)} + q_{ij}^{(2)} = |R_F^2(o_i)| \)

Thus, we have

\[
H^{(1)}(F|d) = - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^1(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^1(o_i) \cap D_j|}{|R_F^1(o_i)|}
= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{d_{ij}^{(1)}}{n} \log_2 \frac{d_{ij}^{(1)}}{d_{ij}^{(1)} + d_{ij}^{(2)}}
\leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{r} f(d_{ij}^{(1)} \cap d_{ij}^{(2)}).
\]

Since \( p_{ij}^{(1)} \leq q_{ij}^{(1)} , p_{ij}^{(2)} \leq q_{ij}^{(2)} \), we have

\[
f(p_{ij}^{(1)} \cap p_{ij}^{(2)}) \leq f(q_{ij}^{(1)} \cap q_{ij}^{(2)}) \leq f(q_{ij}^{(1)} , q_{ij}^{(2)}).
\]

Thus

\[
H^{(1)}(F|d) \leq H^{(2)}(F|d).
\]

**Definition 13:** Suppose that \((O, C \cup \{d\})\) is an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Then \(\lambda\)-joint information entropy of \(F\) and \(d\) is defined as

\[
H^{(\lambda)}(d \cup F) = - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^0(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^0(o_i) \cap D_j|}{|R_F^0(o_i)|}
= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{d_{ij}^{(0)}}{n} \log_2 \frac{d_{ij}^{(0)}}{d_{ij}^{(0)} + d_{ij}^{(1)} + d_{ij}^{(2)}}
\leq \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{r} f(d_{ij}^{(0)} \cap d_{ij}^{(1)} \cap d_{ij}^{(2)}).
\]

Proposition 4: Suppose that \((O, C \cup \{d\})\) is an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Then

\[
H^{(\lambda)}(d|F) = H^{(\lambda)}(d \cup F) - H^{(\lambda)}(F).
\]

Proof: Note that \(\{D_1, D_2, \ldots, D_r\}\) is a partition of \(O\). Then \(\forall i,\)

\[
\sum_{j=1}^{r} |R_F^0(o_i) \cap D_j| = |R_F^0(o_i)|.
\]

Note that \(\{D_1, D_2, \ldots, D_r\}\) is a partition of \(O\). Then \(\forall i,\)

\[
\sum_{j=1}^{r} |R_F^0(o_i) \cap D_j| = |R_F^0(o_i)|.
\]

Thus

\[
H^{(\lambda)}(d|F) = - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^0(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^0(o_i) \cap D_j|}{|R_F^0(o_i)|}
= - \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{d_{ij}^{(0)}}{n} \log_2 \frac{d_{ij}^{(0)}}{n}
\]

By Definition 13, we have

\[
H^{(\lambda)}(F) = - \sum_{i=1}^{n} \frac{|R_F^0(o_i)|}{n} \log_2 \frac{|R_F^0(o_i)|}{n}.
\]

Note that \(\{D_1, D_2, \ldots, D_r\}\) is a partition of \(O\). Then \(\forall i,\)

\[
\sum_{j=1}^{r} |R_F^0(o_i) \cap D_j| = |R_F^0(o_i)|.
\]

Then

\[
H^{(\lambda)}(F) \geq H^{(\lambda)}(d \cup F).
\]

By Proposition 4,

\[
H^{(\lambda)}(d|F) = H^{(\lambda)}(d \cup F) - H^{(\lambda)}(F).
\]

Hence \(H^{(\lambda)}(d|F) \geq 0\).
Lemma 1: Suppose that \((O, C \cup \{d\})\) is an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). If \(R_F^\lambda \subseteq T_d\), then \(\forall o \in O, \forall j,\)

\[
R_F^\lambda(o) \cap D_j = \begin{cases} 
R_F^\lambda(o) & o \in D_j \\
\emptyset & x \notin D_j.
\end{cases}
\]

Proof: If \(o \in D_j\), then \(D_j = T_d(o)\). Since \(R_F^\lambda \subseteq T_d\), we have \(R_F^\lambda(o) \subseteq T_d(o)\). Thus \(R_F^\lambda(o) \cap D_j = R_F^\lambda(o)\).

If \(x \notin D_j\), then \(T_d(o) \cap D_j = \emptyset\). Since \(R_F^\lambda \subseteq T_d\), we have \(R_F^\lambda(o) \subseteq T_d(o)\). Thus \(R_F^\lambda(o) \cap D_j = \emptyset\).

Lemma 2: Suppose that \((O, C \cup \{d\})\) is an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). If \(R_F^\lambda \subseteq T_d\), then \(\forall o \in O,\)

\[
\sum_{j=1}^{r} \frac{|R_F^\lambda(o) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o) \cap D_j|}{n} = \sum_{j=1}^{r} \frac{|R_F^\lambda(o) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o)\|}{n}.
\]

Proof: Note that \(|D_1, D_2, \cdots, D_r\) is a partition of \(O\). Then \(R_F^\lambda(o) \cap D_j = \emptyset\).

Since \(R_F^\lambda \subseteq T_d\), by Lemma 1, we have

\[
R_F^\lambda(o) \cap D_j = \begin{cases} 
R_F^\lambda(o) & j = j^* \\
\emptyset & j \neq j^*.
\end{cases}
\]

Thus

\[
\sum_{j=1}^{r} \frac{|R_F^\lambda(o) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o) \cap D_j|}{n} = \sum_{j=1}^{r} \frac{|R_F^\lambda(o)|}{n} \log_2 \frac{|R_F^\lambda(o)|}{n}.
\]

Theorem 6: Suppose that \((O, C \cup \{d\})\) is an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Then the following conditions are equivalent:

1. \(R_F^\lambda \subseteq T_d\);
2. \(H^\lambda(d \cup F) = H^\lambda(F)\);
3. \(H^\lambda(d | F) = 0\).

Proof: (1) \(\Rightarrow\) (2). This holds by Lemma 2.

(2) \(\Rightarrow\) (1). Suppose \(H^\lambda(d \cup F) = H^\lambda(F)\). Then

\[
-\sum_{i=1}^{n} \frac{|R_F^\lambda(o_i)|}{n} \log_2 \frac{|R_F^\lambda(o_i)|}{n} = -\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o_i) \cap D_j|}{n}.
\]

Since \(|D_1, D_2, \cdots, D_r\) constitutes a partition of \(O\), we can obtain that \(\forall i,\)

\[
\sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} = |R_F^\lambda(o_i)|.
\]

Then

\[
-\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o_i)|}{n} = -\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o_i)|}{n}.
\]

Thus

\[
\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o_i)|}{n} = 0.
\]

This is a contradiction.

Hence \(R_F^\lambda \subseteq T_d\).

(2) \(\Leftrightarrow\) (3). It can be obtained by Proposition 4.

Theorem 7: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(\lambda \in [0, 1]\). Then \((O, C \cup \{d\})\) is \(\lambda\)-consistent \(\iff H^\lambda(d | C) = 0\).

Proof: It can be proved by Theorems 1 and 6.

V. ATTRIBUTE REDUCTION IN AN ISVDIS

In this section, we study attribute reduction in an ISVDIS.

Definition 14: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Then \(F\) is said to be a \(\lambda\)-coordinate subset of \(C\) relative to \(d\), if

\[
POS_F^\lambda(d) = POS_C^\lambda(d).
\]

In this paper, the set of all \(\lambda\)-coordination subsets of \(C\) relative to \(d\) is denoted by \(co^\lambda(C)\).

Definition 15: Let \((O, C \cup \{d\})\) be an ISVDIS. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Then \(F\) is said to be a \(\lambda\)-redut of \(C\) relative to \(d\), if \(F \in co^\lambda(C)\) and \(\forall e \in F, F - \{e\} \notin co^\lambda(C)\).

In this paper, the set of all \(\lambda\)-reduts of \(C\) relative to \(d\) is denoted \(red^\lambda(C)\).

Theorem 8: Suppose that \((O, C \cup \{d\})\) is \(\lambda\)-consistent. Given \(F \subseteq C\) and \(\lambda \in [0, 1]\). Then the following conditions are equivalent:

\[
\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o_i) \cap D_j|}{n} - \log_2 \frac{|R_F^\lambda(o_i) \cap D_j|}{n} = 0.
\]

Thus

\[
\sum_{i=1}^{n} \sum_{j=1}^{r} \frac{|R_F^\lambda(o_i) \cap D_j|}{n} \log_2 \frac{|R_F^\lambda(o_i) \cap D_j|}{n} = 0.
\]

Suppose \(R_F^\lambda \not\subseteq T_d\). Then \(\exists i^*, R_F^\lambda(o_{i^*}) \not\subseteq T_d(o_{i^*})\). Denote \(T_d(o_{i^*}) = D_{i^*}\).

We have

\[
|R_F^\lambda(o_{i^*})| > |R_F^\lambda(o_{i^*}) \cap D_{i^*}|.
\]

It follows that

\[
\frac{|R_F^\lambda(o_{i^*}) \cap D_{i^*}|}{n} \log_2 \frac{|R_F^\lambda(o_{i^*})|}{|R_F^\lambda(o_{i^*}) \cap D_{i^*}|} > 0.
\]

Note that

\[
\forall i, j, |R_F^\lambda(o_i) \cap D_j| \geq |R_F^\lambda(o_j) \cap D_j|.
\]

Then

\[
\forall i, j, |R_F^\lambda(o_i) \cap D_j| \log_2 \frac{|R_F^\lambda(o_i)|}{|R_F^\lambda(o_j) \cap D_j|} \geq 0.
\]

So

\[
\sum_{i=1}^{n} \sum_{j=1}^{r} |R_F^\lambda(o_i) \cap D_j| \log_2 \frac{|R_F^\lambda(o_i)|}{|R_F^\lambda(o_j) \cap D_j|} > 0.
\]

This is a contradiction.

Hence \(R_F^\lambda \subseteq T_d\).
(1) \( F \in co^k(C) \);
(2) \( \Gamma^k_F(d) = \Gamma^0_C(d) \);
(3) \( \delta^k_F = \delta^0_C \);
(4) \( H^k(d|F) = H^0(d|C) \).

Proof: (1) \(\Rightarrow\) (2). Obviously.

(2) \(\Rightarrow\) (3). Suppose \( \Gamma^k_F(d) = \Gamma^0_C(d) \). Then by Theorem 4, we have
\[
\sum_{j=1}^{r} \left( |\underline{R}^k_F(D_j)| - |R^0_C(D_j)| \right) = 0.
\]

By Theorem 2, \( \forall j, \underline{R}^k_F(D_j) \supseteq R^0_C(D_j) \). This implies that \( \forall j, \)
\[
|\underline{R}^k_F(D_j)| - |R^0_C(D_j)| \geq 0.
\]

Then \( \forall j, \)
\[
|\underline{R}^k_F(D_j)| - |R^0_C(D_j)| = 0.
\]

It follows that \( \forall j, \)
\[
\underline{R}^k_F(D_j) = R^0_C(D_j).
\]

Thus \( \forall j, \)
\[
\underline{R}^k_F(o) \subseteq D_j \Leftrightarrow R^0_C(o) \subseteq D_j.
\]

Since \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent, by Theorem 1, we have \( R^0_C \subseteq T_d \). Then \( \forall o \in O, \)
\[
R^0_C(o) \subseteq T_d(o).
\]

Note that \( T_d(o) \in \{D_1, D_2, \cdots, D_r\} \). Then \( \forall o \in O, \)
\[
R^0_C(o) \subseteq T_d(o).
\]

This implies that \( R^0_F \subseteq T_d \).

By Corollary 1,
\[
\forall o \in O, |\Delta^k_F(o)| = 1.
\]

This implies that \( \forall o \in O, |\Delta^0_C(o)| = 1 \).

Thus \( \forall o \in O, \Delta^k_F(o) = \Delta^0_C(o) \).

(3) \(\Rightarrow\) (1). Suppose \( \Delta^k_F = \Delta^0_C \). This implies that \( \forall o \in O, \)
\[
\Delta^0_C(o) = \Delta^0_C(o).
\]

Since \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent, we have \( \forall o \in O, |\Delta^0_C(o)| = 1 \). Then \( \forall o \in O, |\Delta^0_C(o)| = 1 \). By Corollary 1,
\[
R^0_C \subseteq T_d. \quad R^0_F \subseteq T_d.
\]

Suppose that \( \exists j^*, \)
\[
\underline{R}^k_F(D_{j^*}) \not\subseteq R^0_C(D_{j^*}).
\]

Then \( \underline{R}^k_F(D_{j^*}) - R^0_C(D_{j^*}) \neq \emptyset. \)
\[
x^* \in \underline{R}^k_F(D_{j^*}) - R^0_C(D_{j^*}).
\]

It follows that
\[
x^* \in \underline{R}^k_F(D_{j^*}), \quad x^* \not\in R^0_C(D_{j^*}).
\]

\( x^* \in \underline{R}^k_F(D_{j^*}) \) implies that \( x^* \in R^0_C(x^*) \subseteq D_{j^*} \). Then \( D_{j^*} = T_d(x^*) \).

\( x^* \notin \underline{R}^0_C(D_{j^*}) \) implies that \( R^0_C(x^*) \not\subseteq D_{j^*} = T_d(x^*) \). Then \( R^0_C \not\subseteq T_d \). This is a contradiction.

Thus \( \forall j, \)
\[
R^0_C(D_j) \subseteq \underline{R}^0_C(D_j).
\]

By Theorem 2, \( \forall j, R^0_C(D_j) \supseteq R^0_C(D_j). \)
Then \( \forall j, \)
\[
R^0_C(D_j) = R^0_C(D_j).
\]

Hence
\[
POS^k_F(d) = \bigcup_{j=1}^{r} R^0_C(D_j) = \bigcup_{j=1}^{r} R^0_C(D_j) = POS^0_C(d).
\]

(3) \(\Rightarrow\) (4). Suppose \( \delta^k_F = \delta^0_C \). This implies that \( \forall o \in O, \)
\[
\delta^0_C(o) = \delta^0_C(o).
\]

Since \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent, we have \( \forall o \in O, |\Delta^0_C(o)| = 1 \). Then \( \forall o \in O, |\Delta^0_C(o)| = 1 \). By Corollary 1,
\[
R^0_F \subseteq T_d.
\]

By Theorem 6, \( H^k(d|F) = 0 \).
Since \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent, by Theorem 7, we have \( H^k(d|C) = 0 \).

Hence
\[
H^k(d|F) = H^k(d|C).
\]

(4) \(\Rightarrow\) (3). Suppose \( H^k(d|F) = H^k(d|C) \).
Since \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent, by Theorem 7, we have \( H^k(d|C) = 0 \).

Then,
\[
H^k(d|F) = 0.
\]

By Theorem 6,
\[
R^0_F \subseteq T_d.
\]

By Corollary 1, \( \forall o \in O, |\Delta^0_C(o)| = 1 \).
Since \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent, we have \( \forall o \in O, |\Delta^0_C(o)| = 1 \).

This implies that \( \forall o \in O, |\Delta^0_C(o)| = 1 \).

Thus \( \forall o \in O, \Delta^0_C(o) = \Delta^0_C(o) \).

This implies that \( \forall o \in O, \Delta^0_C(o) = \Delta^0_C(o) \).

Thus \( \Delta^0_F = \Delta^0_C \).

\( \square \)

Theorem 9: Assume that \( (O, C \cup \{d\}) \) is \(\lambda\)-consistent. Given \( F \subseteq C \) and \( \lambda \in [0, 1] \). Then the following conditions are equivalent:

(1) \( F \in \text{co}^k(C) \);
(2) \( \Gamma^k_F(d) = \Gamma^0_C(d) \) and \( \forall e \in F, \Gamma^k_{F-e}(d) = \Gamma^0_C(d) \);
(3) \( \delta^k_F = \delta^0_C \) and \( \forall e \in F, \delta^k_{F-e} \neq \delta^0_C \);
(4) \( H^k(d|F) = H^k(d|C) \) and \( \forall e \in F, H^k(d|F - \{e\}) \neq H^k(d|C) \).

Proof: It can be obtained by Proposition 4 and Theorem 8.

Based on the discussion above, we propose the following algorithm for attribute reduction based on similarity measurement and entropy measurement for an ISVDIS.
**VI. AN ILLUSTRATIVE EXAMPLE**

Consider an ISVDIS shown in TABLE 2. ∀ o, o’ ∈ O, ∀ e ∈ E, the similarity degree between e(o) and e(o’), calculated by using Definition 4, is given in TABLES 3-10. ∀ i, ∀ o ∈ O, the λ-tolerance class of o is given in TABLES 11-13. We compute the dependency degree of the decision attribute d over the conditional attribute e1 as follows, taking λ = 0.4. We can easily obtain O/d = {D1, D2, D3}, where D1 = {o1, o6, o8, o110}, D2 = {o2, o5, o9, o111}, D3 = {o4, o7, o12}. Now, λ-conditional entropy of e1 to d is calculated as:

Algorithm 2 Attribute Reduction Algorithm Based on λ-Conditional Entropy for an ISVDIS (CEIS)

**Input:** An ISVDIS (O, C ∪ {d}) and λ ∈ [0, 1].

**Output:** A λ-reduct F.

P ← ∅; Hnew = 1; Hold = 1;

**do**

Temp ← P;

Hnew = Hord;

for each e ∈ (C − P) do

if $H_{e|d}(d) < H_{Temp}(d)$ then

Temp ← P ∪ {e};

Hord = $H_{Temp}(d)$;

P ← Temp;

end

while $H_{new} < H_{old}$;

return F.

$H^λ(d|e_1) = - \sum_{i=1}^{12} \sum_{j=1}^{3} \frac{|R^λ_{e_1}(o_i) \cap D_j|}{n} \log_2 \frac{|R^λ_{e_1}(o_i) \cap D_j|}{|R^λ_{e_1}(o_i)|}$

$= -[(\frac{3}{12} \times \log_2 3 + \frac{4}{12} \times \log_2 4 + \frac{3}{12} \times \log_2 3) + \ldots + (\frac{0}{12} \times \log_2 0 + \frac{0}{12} \times \log_2 0 + \frac{1}{12} \times \log_2 1)]$

$= 6.6930$

Similarly, we could calculate λ-conditional entropies of decision attribute with respect to other conditional attributes:

$H^λ(d|e_2) = 6.0116$, $H^λ(d|e_3) = 6.3465$, $H^λ(d|e_4) = 1.7925$, $H^λ(d|e_5) = 5.9059$, $H^λ(d|e_6) = 6.1658$, $H^λ(d|e_7) = 2.2600$, $H^λ(d|e_8) = 5.8075$.

Since, $H^λ(d|e_4) < H^λ(d|e_7) < H^λ(d|e_8) < H^λ(d|e_5) < H^λ(d|e_2) < H^λ(d|e_6) < H^λ(d|e_3) < H^λ(d|e_1)$, e4 is first reduct member. We will add other attributes to e4 one by one and calculate corresponding λ-conditional entropies. We can easily find the corresponding tolerance classes first, and these results are shown in TABLES 14.

Hence, we have

$H^λ(d|e_1, e_4) = 1.0629$, $H^λ(d|e_2, e_4) = 0.8962$, $H^λ(d|e_3, e_4) = 1.1258$, $H^λ(d|e_4, e_5) = 1.1258$, $H^λ(d|e_4, e_6) = 0.9591$, $H^λ(d|e_4, e_7) = 0.2296$, $H^λ(d|e_4, e_8) = 1.1258$.

Since, $H^λ(d|e_4, e_7) < H^λ(d|e_2, e_4) < H^λ(d|e_4, e_6)$ < $H^λ(d|e_1, e_4) < H^λ(d|e_4, e_5) = H^λ(d|e_4, e_8)$, e7 is second reduct member.

Then, we will add other attributes to {e4, e7} one by one and compute the corresponding tolerance classes (see TABLE 15). And we calculate corresponding λ-conditional entropies as follows:

$H^λ(d|e_1, e_4, e_7) = 0.2296$, $H^λ(d|e_2, e_4, e_7) = 0.1667$, $H^λ(d|e_3, e_4, e_7) = 0.1667$, $H^λ(d|e_4, e_5, e_7) = 0.1667$, $H^λ(d|e_4, e_6, e_7) = 0.1667$, $H^λ(d|e_4, e_7, e_8) = 0.1667$.

Since, $H^λ(d|e_2, e_4, e_7) = H^λ(d|e_3, e_4, e_7) = H^λ(d|e_4, e_5, e_7) = H^λ(d|e_4, e_6, e_7) = H^λ(d|e_4, e_7, e_8) < H^λ(d|e_1, e_4, e_7)$, any one of {e2, e3, e5, e6, e8} will be the member of reduction set. Suppose, e6 is second reduct member.

Similarly, we will add other attributes to {e4, e6, e7} one by one and compute the corresponding tolerance classes (see TABLE 16). Then we calculate corresponding λ-conditional entropies as follows:

$H^λ(d|e_1, e_4, e_6, e_7) = 0.1667$, $H^λ(d|e_2, e_4, e_6, e_7) = 0.1667$, $H^λ(d|e_3, e_4, e_6, e_7) = 0.1667$, $H^λ(d|e_4, e_5, e_6, e_7) = 0.1667$, $H^λ(d|e_4, e_6, e_7, e_8) = 0.0000$.

Since λ-conditional entropy cannot less than 0, {e4, e6, e7, e8} will be the reduct set of an ISVDIS as given in TABLE 2.

**VII. EXPERIMENTAL RESULTS AND ANALYSIS**

This section evaluates the performance of the proposed algorithm and existing methods. The frame work chart of the experiment is displayed in FIGURE 2.

We consider comparing the proposed algorithm with four other algorithms. These are representative attribute reduction algorithms based on fuzzy similarity-based rough set approach (FSRS) [32], fuzzy rough set (FRSM) [9] and dominance relation (DRM) [22]. For the sake of verifying the performance of the proposed attribute reduction algorithm for an ISVDIS, we carry out some experiments on a personal PC with an Intel 3.0-GHz CPU and 64-GB RAM. We execute our experiments on six real datasets selected from the University of California, Irvine (UCI) Machine Learning Repository. These six real datasets are incomplete decision systems (special case of ISVDIS) given in TABLE 17. We take away
TABLE 14. The tolerance class of each object under $R_0^\lambda$ ($P = \{e_1, e_4\}, \{e_2, e_4\}, \{e_3, e_4\}, \{e_4, e_5\}, \{e_4, e_6\}, \{e_4, e_7\}, \{e_4, e_8\}, \lambda = 0.4$).

| $\lambda$ | $R_{e_1,e_4}^\lambda$ | $R_{e_2,e_4}^\lambda$ | $R_{e_3,e_4}^\lambda$ | $R_{e_4,e_5}^\lambda$ | $R_{e_4,e_6}^\lambda$ | $R_{e_4,e_7}^\lambda$ | $R_{e_4,e_8}^\lambda$ |
|-----------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 0.4       | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     |
| 0.36      | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     | $\{e_1, e_4\}$     |

TABLE 15. The tolerance class of each object under $R_0^\lambda$ ($P = \{e_1, e_4, e_7\}, \{e_2, e_4, e_7\}, \{e_3, e_4, e_7\}, \{e_4, e_5, e_7\}, \{e_4, e_6, e_7\}, \{e_4, e_7, e_8\}, \lambda = 0.4$).

| $\lambda$ | $R_{e_1,e_4,e_7}^\lambda$ | $R_{e_2,e_4,e_7}^\lambda$ | $R_{e_3,e_4,e_7}^\lambda$ | $R_{e_4,e_5,e_7}^\lambda$ | $R_{e_4,e_6,e_7}^\lambda$ | $R_{e_4,e_7,e_8}^\lambda$ |
|-----------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| 0.4       | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        |
| 0.36      | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        | $\{e_1, e_4, e_7\}$        |

TABLE 16. The tolerance class of each object under $R_0^\lambda$ ($P = \{e_1, e_4, e_6, e_7\}, \{e_2, e_4, e_6, e_7\}, \{e_3, e_4, e_6, e_7\}, \{e_4, e_5, e_6, e_7\}, \{e_4, e_6, e_7, e_8\}, \lambda = 0.4$).

| $\lambda$ | $R_{e_1,e_4,e_6,e_7}^\lambda$ | $R_{e_2,e_4,e_6,e_7}^\lambda$ | $R_{e_3,e_4,e_6,e_7}^\lambda$ | $R_{e_4,e_5,e_6,e_7}^\lambda$ | $R_{e_4,e_6,e_7,e_8}^\lambda$ |
|-----------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| 0.4       | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   |
| 0.36      | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   | $\{e_1, e_4, e_6, e_7\}$   |

5% information values randomly from the above all data sets in TABLE 17 to obtain ISVDISs. For the experimental work, we use the WEKA tool with ten fold cross-validation technique [17]. To evaluate these attribute reduction approaches, we employ two learning mechanisms to create classifiers. They are frequently-used classifiers which the one is PART, the other is J48.

A. REDUCT SIZE AND CLASSIFICATION ACCURACY

By comparing the selected dataset with the other three methods, the reduced average attribute subset size is given in TABLE 18. It can be observed that compared with the other three approaches, the proposed approach provides smaller or the same reduction size. As for the Soy dataset having 35 attributes, the proposed approach (CEIS) selects 9 attributes while FSRS, FRSM and DRM select 13, 18 and 16 attributes, respectively. The results show that CEIS algorithm has the ability to remove redundant attributes, while other algorithms do not completely remove redundant attributes from the selected attribute subset.

TABLEs 19-20 present the classification results with PART and J48. The underlined symbol expresses the highest classification accuracy among these attribute reduction techniques. Classification accuracies are expressed as percentages. In TABLE 19, for Aud, Soy, Der, Hep and Zoo datasets, classification accuracies evaluated by CEIS algorithm are higher compared to rest three algorithms. For Pro dataset, CEIS executes as well as FSRS in classifier PART while performs better than FRSM and DRM. As can be clearly seen from TABLE 20, for Aud dataset, CEIS is not good enough for DRM in classifier J48. In Pro dataset, classification accuracies evaluated by CEIS algorithm are equal as compared to FSRS algorithm. However, for Soy, Der, Hep and Zoo datasets, the classification accuracies of CEIS do better than other three attribute reduction methods. On the whole, classification accuracies based on the CEIS method were higher than the other three approaches in most case. Therefore, our algorithm is superior to other algorithms. We may conclude that the CEIS method is more effective for attribute reduction in an ISVDIS.

More detailed trend lines of each method on six data sets are shown in FIGUREs 3-5. FIGURE 3 shows the reduction comparison of the average attribute subset size of all four algorithms on these six datasets. It can be seen that the CEIS algorithm proposed in this article selects the least attributes as the elements of the reduction set. FIGUREs 4-5 show the more detailed trend of the algorithm classification accuracy with the number of selected attributes on all selected data sets.
It can be seen from the figure that the proposed method provides higher or almost equal classification accuracy for all six data sets. By comparing the above tables and charts, we can conclude that the proposed CEIS algorithm is an acceptable method to select the optimal attribute subset in an ISVDIS.
TABLE 18. The reduct of different \( \lambda \) value for 4 algorithms.

| \( \lambda \) | Algorithm | Number |
|---|---|---|
| | | Adult | APS | Audi | Derm | Hepa | Proc | Stat | Soyb | Weat | Zoo |
| 0.1 | FSRS | 2 | 4 | 14 | 7 | 3 | 3 | 6 | 13 | 4 | 5 |
| | FRSM | 5 | 13 | 34 | 19 | 9 | 6 | 12 | 19 | 10 | 13 |
| | DRM | 5 | 10 | 35 | 12 | 12 | 5 | 10 | 18 | 11 | 9 |
| | CEIS | 2 | 3 | 16 | 6 | 2 | 3 | 5 | 17 | 4 | 7 |
| 0.3 | FSRS | 2 | 4 | 14 | 7 | 3 | 3 | 6 | 13 | 4 | 5 |
| | FRSM | 5 | 13 | 35 | 19 | 9 | 6 | 12 | 28 | 8 | 13 |
| | DRM | 4 | 10 | 37 | 12 | 7 | 5 | 12 | 20 | 9 | 9 |
| | CEIS | 2 | 4 | 6 | 6 | 3 | 3 | 9 | 9 | 3 | 6 |
| 0.5 | FSRS | 2 | 4 | 14 | 7 | 3 | 3 | 6 | 13 | 4 | 5 |
| | FRSM | 5 | 13 | 35 | 19 | 9 | 6 | 13 | 18 | 9 | 9 |
| | DRM | 4 | 10 | 32 | 12 | 7 | 7 | 12 | 18 | 9 | 9 |
| | CEIS | 2 | 3 | 6 | 6 | 2 | 3 | 5 | 9 | 4 | 6 |
| 0.7 | FSRS | 2 | 4 | 14 | 7 | 3 | 3 | 6 | 13 | 4 | 5 |
| | FRSM | 5 | 13 | 35 | 19 | 9 | 6 | 13 | 18 | 9 | 9 |
| | DRM | 4 | 10 | 32 | 12 | 7 | 7 | 12 | 18 | 9 | 9 |
| | CEIS | 2 | 3 | 6 | 6 | 2 | 3 | 5 | 9 | 4 | 6 |
| 0.9 | FSRS | 2 | 4 | 14 | 7 | 3 | 3 | 6 | 13 | 4 | 5 |
| | FRSM | 5 | 13 | 34 | 19 | 9 | 6 | 13 | 18 | 9 | 9 |
| | DRM | 4 | 9 | 33 | 12 | 7 | 7 | 12 | 18 | 9 | 9 |
| | CEIS | 2 | 4 | 2 | 5 | 3 | 3 | 9 | 9 | 4 | 6 |

TABLE 19. Comparison of classification accuracies (rules-PART).

| \( \lambda \) | Algorithm | Accuracies |
|---|---|---|
| | | Adult | APS | Audi | Derm | Hepa | Proc | Stat | Soyb | Weat | Zoo |
| 0.1 | FSRS | 76.80 | 75.49 | 75.66 | 80.33 | 79.35 | 55.45 | 83.95 | 83.06 | 73.44 | 97.03 |
| | FRSM | 82.30 | 91.90 | 76.11 | 93.17 | 77.42 | 51.49 | 84.20 | 68.73 | 73.50 | 91.09 |
| | DRM | 81.90 | 87.99 | 76.99 | 95.08 | 78.63 | 49.84 | 82.80 | 79.80 | 73.44 | 97.41 |
| | CEIS | 82.42 | 92.13 | 77.43 | 95.09 | 79.35 | 55.45 | 85.70 | 82.41 | 73.44 | 97.41 |
| 0.3 | FSRS | 76.80 | 70.49 | 76.54 | 80.33 | 78.06 | 45.51 | 83.10 | 83.06 | 73.44 | 97.03 |
| | FRSM | 82.10 | 91.90 | 76.55 | 93.17 | 77.42 | 51.49 | 84.00 | 68.73 | 73.67 | 91.09 |
| | DRM | 81.00 | 87.99 | 76.11 | 95.08 | 78.01 | 49.84 | 82.80 | 79.80 | 73.44 | 97.91 |
| | CEIS | 83.13 | 92.10 | 76.48 | 95.10 | 78.61 | 54.13 | 83.30 | 83.12 | 77.06 | 97.21 |
| 0.5 | FSRS | 76.80 | 69.18 | 76.54 | 80.33 | 78.06 | 52.48 | 83.95 | 83.06 | 73.44 | 97.03 |
| | FRSM | 83.60 | 91.90 | 77.88 | 92.90 | 77.42 | 51.49 | 84.30 | 68.73 | 73.57 | 91.09 |
| | DRM | 81.00 | 87.44 | 69.91 | 94.54 | 77.29 | 55.16 | 82.80 | 79.15 | 73.44 | 91.09 |
| | CEIS | 84.07 | 92.11 | 77.98 | 95.51 | 78.42 | 56.16 | 85.85 | 83.24 | 96.04 | 97.07 |
| 0.7 | FSRS | 76.80 | 73.99 | 78.76 | 94.53 | 67.12 | 53.79 | 83.65 | 83.06 | 73.44 | 92.07 |
| | FRSM | 82.36 | 91.90 | 73.89 | 82.51 | 78.06 | 56.03 | 83.05 | 88.57 | 76.67 | 96.03 |
| | DRM | 81.18 | 87.99 | 74.33 | 92.89 | 80.00 | 53.06 | 83.95 | 85.21 | 73.44 | 93.06 |
| | CEIS | 83.01 | 91.93 | 78.77 | 95.16 | 80.64 | 56.13 | 83.97 | 92.08 | 82.41 | 96.09 |
| 0.9 | FSRS | 76.80 | 70.33 | 75.22 | 83.06 | 78.06 | 50.17 | 83.65 | 83.06 | 73.44 | 97.03 |
| | FRSM | 81.90 | 91.90 | 76.55 | 94.81 | 77.42 | 53.47 | 83.05 | 68.73 | 73.67 | 91.09 |
| | DRM | 81.00 | 87.94 | 76.11 | 94.81 | 77.93 | 55.16 | 83.95 | 82.41 | 74.33 | 93.07 |
| | CEIS | 81.1 | 92.01 | 76.6 | 94.82 | 78.09 | 55.21 | 83.96 | 83.07 | 73.71 | 97.1 |

B. FRIEDMAN TEST AND NEMENYI TEST

Based on the obtained classification accuracies in the previous section, these two tests are used to further verify the stability of the proposed approach in this part.

Friedman test, as a nonparametric test method in statistics, is used to compare the overall performance of \( k \) algorithms on \( N \) data sets, and to draw the conclusion whether there are differences in the performance of these algorithms. If there are differences, we will carry out the next more detailed test, namely Nemenyi test. The Nemenyi test is used for determining which algorithms differ statistically in performance.

Suppose that \( N \) and \( k \) express the number of data sets and algorithms, respectively. \( \forall \ i \), \( r_i \) can be viewed as the mean ordering of the \( i \)-th algorithm. Friedman statistic, denoted by \( F_F \), is defined as

\[
F_F = \frac{(N - 1)\chi^2_F}{N(k-1) - \chi^2_F},
\]

where

\[
\chi^2_F = \frac{12N}{k(k+1)} \sum_{i=1}^{k} r_i^2 - 3N(k+1).
\]

This statistic \( F_F \) based on the F-distribution and distributed by \( k - 1 \) and \((k - 1)(N - 1)\) degrees of freedom. In Friedman test, if the value of \( F_F \) calculated is greater than \( F_{\alpha}(k - 1, N - 1) \), then the original hypothesis does not hold. Nemenyi test can be used to inquire into which algorithm
TABLE 20. Comparison of classification accuracies (trees-J48).

| λ    | Algorithm | Adult | APS | Audi | Derm | Hepa | Proc | Stat | Soyb | Weat | Zoo |
|------|-----------|-------|-----|------|------|------|------|------|------|------|-----|
| 0.1  | FSRS      | 78.80 | 75.54 | 72.56 | 80.33 | 79.35 | 55.45 | 77.95 | 89.58 | 71.67 | 94.06 |
|      | FRSM      | 82.70 | 89.09 | 72.12 | 93.39 | 81.29 | 55.78 | 80.90 | 64.50 | 76.72 | 96.04 |
|      | DRM       | 82.30 | 84.74 | 71.23 | 94.54 | 79.35 | 53.47 | 79.20 | 85.34 | 77.61 | 94.06 |
|      | CEIS      | 83.01 | 90.05 | 72.59 | 94.71 | 81.44 | 55.89 | 90.97 | 90.45 | 78.11 | 96.49 |
| 0.3  | FSRS      | 78.80 | 75.54 | 72.12 | 80.33 | 79.35 | 54.13 | 80.40 | 77.95 | 71.67 | 94.06 |
|      | FRSM      | 81.90 | 89.09 | 69.03 | 93.99 | 81.29 | 55.78 | 81.00 | 77.83 | 76.94 | 96.04 |
|      | DRM       | 77.90 | 84.04 | 71.24 | 94.54 | 80.03 | 53.47 | 79.20 | 79.20 | 76.89 | 94.06 |
|      | CEIS      | 81.91 | 89.13 | 73.82 | 94.79 | 86.11 | 55.78 | 83.09 | 81.43 | 78.09 | 96.04 |
| 0.5  | FSRS      | 78.80 | 71.59 | 77.87 | 80.33 | 77.42 | 53.47 | 77.95 | 85.58 | 71.67 | 94.06 |
|      | FRSM      | 84.70 | 89.09 | 71.24 | 94.54 | 81.29 | 55.78 | 79.85 | 64.50 | 76.17 | 96.04 |
|      | DRM       | 77.90 | 84.49 | 69.03 | 95.36 | 81.30 | 55.78 | 79.20 | 89.25 | 76.89 | 94.06 |
|      | CEIS      | 84.91 | 90.01 | 77.89 | 95.47 | 81.31 | 64.03 | 84.47 | 91.11 | 76.94 | 93.73 |
| 0.7  | FSRS      | 78.80 | 80.49 | 70.80 | 89.07 | 77.42 | 52.48 | 78.50 | 85.58 | 71.67 | 94.06 |
|      | FRSM      | 83.10 | 89.09 | 72.12 | 97.27 | 81.29 | 52.15 | 80.80 | 64.50 | 77.44 | 96.04 |
|      | DRM       | 77.90 | 84.64 | 71.24 | 93.72 | 81.29 | 55.78 | 79.75 | 89.90 | 77.06 | 93.06 |
|      | CEIS      | 85.07 | 89.93 | 76.99 | 98.41 | 83.82 | 55.62 | 81.09 | 90.08 | 81.53 | 96.77 |
| 0.9  | FSRS      | 78.80 | 71.34 | 72.57 | 83.88 | 77.42 | 53.14 | 78.50 | 85.58 | 71.67 | 94.06 |
|      | FRSM      | 81.90 | 89.09 | 69.91 | 97.27 | 81.29 | 52.15 | 79.45 | 64.50 | 77.44 | 96.04 |
|      | DRM       | 77.90 | 84.29 | 71.24 | 94.54 | 81.29 | 55.78 | 81.45 | 89.90 | 77.39 | 96.04 |
|      | CEIS      | 78.00 | 89.96 | 74.08 | 97.31 | 83.94 | 61.49 | 81.92 | 89.91 | 77.52 | 96.09 |

FIGURE 5. Reduction attribute subset size change of four algorithms when $\lambda = 0.5$.  
FIGURE 6. Reduction attribute subset size change of four algorithms when $\lambda = 0.7$.  
FIGURE 7. Reduction attribute subset size change of four algorithms when $\lambda = 0.9$.  
FIGURE 8. The classification accuracy of the four algorithms varies with the classifier PART when $\lambda = 0.1$.  

is better. Critical difference, denoted as $CD_\alpha$, is defined as follows:

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k + 1)}{6N}},$$

where $q_\alpha$ and $\alpha$ are the critical value and the significance level of this test, respectively. The difference between the average ranking of each pair of algorithms is compared with $CD_\alpha$, and if it is greater than $CD_\alpha$, In other words, the algorithm
with higher average ranking is statistically superior to the algorithm with lower average ranking; otherwise, there is no statistically difference between the two.

Below, we select all datasets in TABLE 17. Friedman test is used to test the influence of parameter $\lambda$ on the classification accuracy of different algorithms, and the test results show parameter $\lambda$ has no significant effect on the classification accuracy of different algorithms. Thus we choose the classification accuracies of $\lambda = 0.1$ to test whether the classification accuracies of the four algorithms are significantly different. From TABLEs 19-20, we can obtain the ranking of classification accuracies of reduced data with PART and J48, respectively (see TABLEs 21-22). Note that $k - 1 = 3$, $(k - 1)(N - 1) = 27$ and $F_{0.05}(3, 27) = 2.96$. Then $F_{F} = 4.344$ with PART and $F_{F} = 20.978$ with J48. Therefore, the two value of $F_{F}$ are outweigh the value of $F_{0.05}(3, 27)$. That is to say, they rejects the original hypothesis at $\alpha = 0.05$ under the Friedman tests. Therefore, there are
statistically significant differences in classification accuracy among these algorithms. Next, based on the Nemenyi test, we can easily figure out that \( q_{\alpha} = 2.3437 \) and \( CD_{\alpha} = 2.3437 \times \sqrt{\frac{4 \times (4+1)}{36}} = 0.17464 \) at \( \alpha = 0.05 \). The results of the Nemenyi test for these algorithms at \( \alpha = 0.05 \) is displays in FIGUREs 18-19. One can conclude that the measures connected with a line are not significantly different. From

FIGURE 15. The classification accuracy of the four algorithms varies with the classifier J48 when \( \lambda = 0.5 \).

FIGURE 16. The classification accuracy of the four algorithms varies with the classifier J48 when \( \lambda = 0.7 \).

FIGURE 17. The classification accuracy of the four algorithms varies with the classifier J48 when \( \lambda = 0.9 \).

FIGURE 18. Nemenyi test with Part.

FIGURE 19. Nemenyi test with J48.

FIGUREs 18-19, we can come to the same conclusions as follows: the classification accuracy of CEIS is significantly higher than that of FSRS and DRM. There is no significant
difference between CEIS and FRSM. There is no significant difference among FSRS, FRSM and DRM.

VIII. CONCLUSION

This article has studied attribute reduction for an ISVDIS by using entropy measurement. Connections between the proposed attribute reduction and λ-information entropy have been researched. An attribute reduction algorithm in an ISVDIS based on information entropy has been proposed. Moreover, we have showed that λ-reduction obtained from λ-information entropy are equivalent to those obtained from λ-rough approximations. Soon after, we will investigate some applications of λ-reduction in an ISVDIS.

THE APPENDIX OF SYMBOLS

For convenience, we give the appendix of these symbols as follows (see TABLE 23).

ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions, which have helped immensely in improving the quality of the paper.

REFERENCES

[1] J. Błaszczyński, R. Słowiński, and M. Szeląg, “Sequential covering rule induction algorithm for variable consistency rough set approaches,” Inf. Sci., vol. 181, no. 5, pp. 987–1002, Mar. 2011.
[2] C. Cornelis, R. Jensen, G. Hurtado, and D. Śleczak, “Attribute selection with fuzzy decision reducts,” Inf. Sci., vol. 180, no. 2, pp. 209–224, Jan. 2010.
[3] Z. C. Chen and K. Y. Qin, “Attribute reduction of set-valued information systems based on a tolerance relation,” Comput. Syst. Sci., vol. 23, no. 1, pp. 18–22, 2010.
[4] I. Couso and D. Dubois, “Statistical reasoning with set-valued information: Ontic vs. epistemic views,” Int. J. Approx. Reasoning, vol. 55, no. 7, pp. 1502–1518, Oct. 2014.
[5] L. Chen, S. Liao, N. Xie, Z. Li, G. Zhang, and C.-F. Wen, “Measures of uncertainty for an incomplete set-valued information system with the optimal selection of subsystems: Gaussian kernel method,” IEEE Access, vol. 8, pp. 21022–21035, 2020.
[6] I. Düntsch and G. Gediga, “Uncertainty measures of rough set prediction,” Artif. Intell., vol. 106, no. 1, pp. 109–137, Nov. 1998.
[7] J. Dai, H. Hu, W.-Z. Wu, Y. Qian, and D. Huang, “Maximal-discriminability-pair-based approach to attribute reduction in fuzzy rough sets,” IEEE Trans. Fuzzy Syst., vol. 26, no. 4, pp. 2174–2187, Aug. 2018.
[8] J. Dai and H. Tian, “Entropy measures and granularity measures for set-valued information systems,” Inf. Sci., vol. 240, pp. 72–82, Aug. 2013.
[9] J. Dai and H. Tian, “Fuzzy rough set model for set-valued data,” Fuzzy Sets Syst., vol. 229, pp. 54–68, Oct. 2013.
[10] D. Dubois and H. Prade, “Rough fuzzy sets and fuzzy rough sets,” Int. J. Gen. Syst., vol. 17, nos. 2-3, pp. 191–209, 1990.
[11] J. Dai, W. Wang, and J.-S. Mi, “Uncertainty measurement for interval-valued information systems,” Inf. Sci., vol. 251, pp. 63–78, Dec. 2013.
[12] 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.
[13] M. Friedman, “A comparison of alternative tests of significance for the problem of M rankings,” Ann. Math. Statist., vol. 11, no. 1, pp. 86–92, Apr. 1940.
[14] 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.
[15] Y. Y. Guan, P. J. Xue, and H. Q. Hu, “Attribute reduction and definite decision rules optimization in set-valued decision information systems,” Syst. Eng. Electron., vol. 28, no. 4, pp. 551–555, 2006.
[16] Q. Hu, W. Pedrycz, D. Yu, and J. Lang, “Selecting discrete and continuous features based on neighborhood decision error minimization,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 40, no. 1, pp. 137–150, Feb. 2010.
[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: An update,” ACM SIGKDD Explor. Newslett., vol. 11, no. 1, pp. 10–18, 2009.
[18] Y. Huang, T. Li, C. Luo, H. Fujita, and S.-J. Horng, “Dynamic variable precision rough set approach for probabilistic set-valued information systems,” Knowl.-Based Syst., vol. 122, pp. 131–147, Apr. 2017.
[19] M. Kryszkiewicz, “Rules in incomplete information systems,” Inf. Sci., vol. 113, nos. 3–4, pp. 271–292, Feb. 1999.
[20] Y. Leung, M. M. Fischer, W.-Z. Wu, and J.-S. Mi, “A rough set approach for the discovery of classification rules in interval-valued information systems,” Int. J. Approx. Reasoning, vol. 47, no. 2, pp. 233–246, Feb. 2008.
[21] G. Lang, Q. Li, and T. Yang, “An incremental approach to attribute reduction of dynamic set-valued information systems,” Int. J. Mach. Learn. Cybern., vol. 5, no. 5, pp. 775–788, Oct. 2014.
[22] Y. Liu and C. Zhong, “Attribute reduction of set-valued decision information system based on dominance relation,” J. Intercdiscipl. Math., vol. 19, no. 3, pp. 469–479, May 2016.
[23] J.-S. Mi, Y. Leung, and W.-Z. Wu, “An uncertainty measure in partition-based fuzzy rough sets,” Int. J. Gen. Syst., vol. 34, no. 1, pp. 77–90, Feb. 2005.
[24] Y. Nakahara, “User oriented ranking criteria and its application to fuzzy mathematical programming problems,” Fuzzy Sets Syst., vol. 94, no. 3, pp. 275–286, Mar. 1998.
[25] Y. Nakahara, M. Sasaki, and M. Gen, “On the linear programming problems with interval coefficients,” Fuzzy Sets Syst., vol. 23, pp. 301–304, Nov. 1992.
[26] Z. Pawlak, “Rough sets,” Int. J. Comput. Inf. Sci., vol. 11, no. 5, pp. 341–356, 1982.
[27] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data. Dordrecht, The Netherlands: Kluwer, 1991.
[28] Y. H. Qian, J. Y. Liang, P. Song, and C. Y. Dang, “On dominance relations in disjunctive set-valued ordered information systems,” Int. J. Inform. Technol. Decis. Making, vol. 9, no. 1, pp. 9–33, Jan. 2010.
[29] Y. Qian, C. Dang, J. Liang, and D. Tang, “Set-valued ordered information systems,” Inf. Sci., vol. 179, no. 16, pp. 2809–2832, Jul. 2009.
[30] H. Sakai, M. Nakata, and D. Slezak, “A prototype system for rule generation in Lipsk’s incomplete information databases,” in Proc. 13th Rough Sets, Fuzzy Sets, Data Mining Granular Comput., 2011, pp. 175–182.
[31] X.-X. Song and W. X. Zhang, “Knowledge reduction in set-valued decision information system,” in Proc. Rough Sets Current Trends Comput. (RSCTC), Uppsala, Sweden, 2009, vol. 7260, no. 1, pp. 348–357.
[32] 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.

[33] K. Thangavel and A. Pethalakshmi, “Dimensionality reduction based on rough set theory: A review,” *Appl. Soft Comput.*, vol. 9, no. 1, pp. 1–12, Jan. 2009.

[34] T. Li, W. Yan, and M. O. Zhi-Wen, “Knowledge reduction in set-valued incomplete information system,” *J. Sichuan Normal Univ.*, vol. 30, no. 3, pp. 288–290, 2007.

[35] M. J. Wierman, “Measuring uncertainty in rough set theory,” *Int. J. Gen. Syst.*, vol. 28, nos. 4–5, pp. 283–297, Oct. 1999.

[36] H. Wang and R. Gao, “Knowledge reduction of set-valued decision information systems based on tolerance relation,” *Appl. Mech. Mater.*, vols. 462–463, pp. 466–471, Nov. 2013.

[37] C. Wang, Y. Huang, W. Ding, and Z. Cao, “Attribute reduction with fuzzy rough self-information measures,” *Inf. Sci.*, vol. 549, pp. 68–86, Mar. 2021.

[38] C. Wang, Y. Huang, M. Shao, and X. Fan, “Fuzzy rough set-based attribute reduction using distance measures,” *Knowl.-Based Syst.*, vol. 164, pp. 205–212, Jan. 2019.

[39] 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.

[40] 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.

[41] H. Wang and H.-B. Yue, “Entropy measures and granularity measures for interval and set-valued information systems,” *Soft Comput.*, vol. 20, no. 9, pp. 3489–3495, Sep. 2016.

[42] N. Xie, M. Liu, Z. Li, and G. Zhang, “New measures of uncertainty for an interval-valued information system,” *Inf. Sci.*, vol. 470, pp. 156–174, Jan. 2019.

[43] Y. Y. Yao and N. Noroozi, “A unified framework for set-based computations,” in *Proc. 3rd Int. Workshop Rough Sets Soft Comput.*, 1994, pp. 10–12.

[44] Y. Y. Yao and X. Li, “Comparison of rough-set and interval-set models for uncertain reasoning,” *Fundam. Informat.*, vol. 27, no. 2,3, pp. 289–298, 1996.

[45] X. Yang, D. Yu, J. Yang, and L. Wei, “Dominance-based rough set approach to incomplete interval-valued information system,” *Data Knowl. Eng.*, vol. 68, no. 11, pp. 1331–1347, Nov. 2009.

[46] X. Zhang, C. Mei, D. Chen, and J. Li, “Multi-confidence rule acquisition and confidence-preserved attribute reduction in interval-valued decision systems,” *Int. J. Approx. Reasoning*, vol. 55, no. 8, pp. 1787–1804, Nov. 2014.

**YUANXIA ZHANG** received the M.Sc. degree from the Department of Mathematics, Dalian University of Technology, Dalian, China, in 2009. He is currently an Associate Professor with the School of Computer Science and Engineering, Yulin Normal University, Yulin, China. His current research interests include image processing, rough set theory, and artificial intelligence.

**ZUOZAN CHEN** received the M.Sc. degree from the Department of Mathematics, Dalian University of Technology, Dalian, China, in 2009. He is currently an Associate Professor with the School of Computer Science and Engineering, Yulin Normal University, Yulin, China. His current research interests include big data, rough set theory, and granular computing.

* * *