

Frontiers of Information Technology & Electronic Engineering
 www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
 ISSN 2095-9184 (print); ISSN 2095-9230 (online)
 E-mail: jzus@zju.edu.cn



Attribute reduction in interval-valued information systems based on information entropies*

Jian-hua DAI^{†1,2}, Hu HU², Guo-jie ZHENG², Qing-hua HU¹, Hui-feng HAN², Hong SHI¹

(¹School of Computer Science and Technology, Tianjin University, Tianjin 300350, China)

(²College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

[†]E-mail: david.joshua@qq.com

Received Dec. 13, 2015; Revision accepted Mar. 4, 2016; Crosschecked Aug. 8, 2016

Abstract: Interval-valued data appear as a way to represent the uncertainty affecting the observed values. Dealing with interval-valued information systems is helpful to generalize the applications of rough set theory. Attribute reduction is a key issue in analysis of interval-valued data. Existing attribute reduction methods for single-valued data are unsuitable for interval-valued data. So far, there have been few studies on attribute reduction methods for interval-valued data. In this paper, we propose a framework for attribute reduction in interval-valued data from the viewpoint of information theory. Some information theory concepts, including entropy, conditional entropy, and joint entropy, are given in interval-valued information systems. Based on these concepts, we provide an information theory view for attribute reduction in interval-valued information systems. Consequently, attribute reduction algorithms are proposed. Experiments show that the proposed framework is effective for attribute reduction in interval-valued information systems.

Key words: Rough set theory, Interval-valued data, Attribute reduction, Entropy
<http://dx.doi.org/10.1631/FITEE.1500447>

CLC number: TP18

1 Introduction

Rough set theory is as an extension of set theory for the study of intelligent systems characterized by insufficient and incomplete information (Pawlak, 1991). It has attracted the attention of many researchers who have studied its theories and applications in recent years (Dai and Xu, 2012; 2013; Dai et al., 2012; 2013a; 2013c; Hu, 2015; Lin et al., 2015; Zhang X et al., 2015; Zhang XH et al., 2016).

The classical rough set model is not appropriate for handling interval-valued data. However, in real applications, many data are interval-valued (Bil-

lard et al., 2008; Hedjazi et al., 2011; Dai et al., 2012; 2013b). Thus, dealing with interval-valued information systems becomes an interesting problem in rough set theory (Leung et al., 2008; Qian et al., 2008; Yang et al., 2009). Dai (2008) investigated the algebraic structures for interval-set-valued rough sets generated from an approximation space. Recently, Dai et al. (2012; 2013b) studied the uncertainty measurement issue in interval-valued information systems. Leung et al. (2008) presented a rough set approach based on the misclassification rate for interval-valued information systems. Qian et al. (2008) proposed a dominance relation to ordered interval information systems. Yang et al. (2009) investigated a dominance relation in incomplete interval-valued information systems. However, there have been few studies dealing with the attribute reduction issue in interval-

[†] Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61473259, 61502335, 61070074, and 60703038), the Zhejiang Provincial Natural Science Foundation (No. Y14F020118), and the PEIYANG Young Scholars Program of Tianjin University, China (No. 2016XRX-0001)

ORCID: Jian-hua DAI, <http://orcid.org/0000-0003-1459-0833>
 © Zhejiang University and Springer-Verlag Berlin Heidelberg 2016

valued information systems. Dai *et al.* (2013b) considered this issue from the viewpoint of uncertainty measurement. In this study, we aim to introduce a heuristic approach for attribute reduction in interval-valued information systems based on information theory.

2 Basic concepts

2.1 Similarity degree between interval values

Unlike classical real values, it is difficult to compare two interval values using traditional methods. Motivated by this fact, some research efforts have been directed toward finding efficient methods to measure or rank two interval values, mainly in the fuzzy set community (Bustince *et al.*, 2006; Zhang and Fu, 2006; Galar *et al.*, 2011). Galar *et al.* (2011) defined the similarity between two intervals as an interval. Zhang and Fu (2006) defined the similarity between two intervals as a real number. In this study, motivated by a similarity measure for interval-valued fuzzy sets proposed by Zhang and Fu (2006), we give a similarity measure for general interval-valued data:

Definition 1 Let $U = \{u_1, u_2, \dots, u_n\}$ be the universe of the interval values, and $u_i = [u_i^-, u_i^+]$, $i = 1, 2, \dots, n$. Let $m^- = \min_{u_i \in U} \{u_i^-\}$, $m^+ = \max_{u_i \in U} \{u_i^+\}$. The relative bound difference similarity degree between u_i and u_j is defined as

$$v_{ij} = 1 - \frac{1}{2} \frac{|u_i^- - u_j^-| + |u_i^+ - u_j^+|}{m^+ - m^-}. \quad (1)$$

Proposition 1 Note that the relative bound difference similarity degree v_{ij} has the following properties:

1. $0 \leq v_{ij} \leq 1$;
2. $v_{ij} = 1$ if and only if u_i equals u_j ;
3. $v_{ij} = v_{ji}$.

Proof It can be proved easily according to Definition 1.

Example 1 Assume $U = \{u_1, u_2\}$, $u_1 = [3, 4]$, $u_2 = [2, 5]$. We have

$$\begin{aligned} m^- &= \min\{3, 2\} = 2, \quad m^+ = \max\{4, 5\} = 5, \\ |u_1^- - u_2^-| &= |3 - 2| = 1, \quad |u_1^+ - u_2^+| = |4 - 5| = 1, \\ v_{ij} &= 1 - \frac{1}{2} \times \frac{1 + 1}{5 - 2} = \frac{2}{3}. \end{aligned}$$

2.2 Similarity classes and generalized decisions in interval-valued information systems

Let $IVIS = (U, A)$ denote an interval-valued information system, where $U = \{u_1, u_2, \dots, u_n\}$ is a non-empty finite set called the universe of discourse, and $A = \{a_1, a_2, \dots, a_m\}$ is a non-empty finite set of m attributes called conditional attributes.

Definition 2 Assume that $IVIS = (U, A)$ is an interval-valued information system. For a given similarity rate $\alpha \in [0, 1]$ and an attribute subset $B \subseteq A$, an α -similarity class of an object $u_i \in U$ is denoted as

$$S_B^\alpha(u_i) = \{u_j \mid v_{ij}^\kappa > \alpha, \forall a_\kappa \in B, u_j \in U\}, \quad (2)$$

where v_{ij}^κ represents the similarity degree of u_i and u_j at the κ th attribute.

Remark 1 S_B^α denotes the family set $\{S_B^\alpha(u_i) \mid u_i \in U\}$.

$S_B^\alpha(u_i)$ is the maximum set of objects which are possibly indiscernible with object u_i by attribute set B under similarity rate α . In other words, $S_B^\alpha(u_i)$ is an α -similarity class of u_i .

Proposition 2 Given an interval-valued information system $IVIS$, and assuming that the attribute subset $B \subseteq A$, then S_B^α has the following properties for any $u_i \in U$:

1. $S_B^\alpha(u_i) = \bigcap_{b \in B} S_{\{b\}}^\alpha(u_i)$;
2. if $C \subseteq B$, then $S_B^\alpha(u_i) \subseteq S_C^\alpha(u_i)$.

Proof

1. By definition, we have $S_B^\alpha(u_i) = \{u_j \mid v_{ij}^b > \alpha, \forall b \in B, u_j \in U\} = \bigcap_{b \in B} S_{\{b\}}^\alpha(u_i)$.

2. By definition, we have $S_B^\alpha(u_i) = \{u_j \mid v_{ij}^b > \alpha, \forall b \in B, u_j \in U\}$ and $S_C^\alpha(u_i) = \{u_j \mid v_{ij}^b > \alpha, \forall b \in C, u_j \in U\}$. For any $u_j \in S_B^\alpha(u_i)$, we know $v_{ij}^b > \alpha$ holds for any $b \in B$. Since $C \subseteq B$, we know $v_{ij}^b > \alpha$ holds for any $b \in C$. Hence, $u_j \in S_C^\alpha(u_i)$.

Let $IVDS = (U, A \cup \{d\})$ denote an interval-valued decision system where d is the decision attribute, also called the class label.

Definition 3 Assume that $IVDS = (U, A \cup \{d\})$ is an interval-valued decision system. The decision class of an object $u_i \in U$ is denoted as

$$D(u_i) = \{u_j \mid d(u_i) = d(u_j), \forall u_j \in U\}. \quad (3)$$

Remark 2 We use $U/d = \{D_i \mid d(u_x) = d(u_y), \forall u_x, u_y \in D_i\}$ to denote the partition of U based on

the decision attribute d ; i.e., $D(u_i)$ represents the set of objects which contain the same decision attribute as u_i .

Similar to the definition of generalized decision proposed by Kryszkiewicz (1998) for incomplete decision systems, α -generalized decision can be defined as follows:

Definition 4 The α -generalized decision of an object $u_i \in U$ is denoted as

$$\partial_B^\alpha(u_i) = \{d(u_j) \mid v_{ij}^\kappa > \alpha, \forall a_\kappa \in B, u_j \in U\}. \quad (4)$$

Remark 3 ∂_B^α denotes the family set $\{\partial_B^\alpha(u_i) \mid u_i \in U\}$.

Let $IVDS = (U, A \cup \{d\})$ be an interval-valued decision system. If $S_A^\alpha(u_i) \subseteq D(u_i)$ for any object $u_i \in U$, then the interval-valued decision system $\{U, A \cup \{d\}\}$ is called a consistent (deterministic, definite) interval-valued decision system. Otherwise, it is called an inconsistent (non-deterministic, non-definite) interval-valued decision system.

Proposition 3 Given a consistent decision system $IVDS = (U, A \cup \{d\})$, for an attribute set $B \subseteq A$, the following conditions are equivalent for all objects $u_i \in U$:

1. $S_B^\alpha(u_i) \subseteq D(u_i)$;
2. $|\partial_B^\alpha(u_i)| = 1$.

Proof Suppose $S_B^\alpha(u_i) \subseteq D(u_i)$. For any $u_j \in S_B^\alpha(u_i)$, we have $u_j \in D(u_i)$. It follows that $d(u_j) = d(u_i)$. In other words, $d(u_j) = d(u_i)$ holds for any u_j satisfying $v_{ij}^b > \alpha, \forall b \in B$. Hence, we know $|\partial_B^\alpha(u_i)| = 1$ by Definition 4.

Suppose $|\partial_B^\alpha(u_i)| = 1$. For any $u_j \in S_B^\alpha(u_i)$, we have $d(u_i) = d(u_j)$. It means $u_j \in D(u_i)$. Thus, we have $S_B^\alpha(u_i) \subseteq D(u_i)$.

3 Information entropy and conditional entropy for interval-valued information systems

Let us introduce an information measure for the discernibility power of an α -similarity class denoted as $S_B^\alpha(u_i)$, which is equivalent to Shannon's entropy if $S_B^\alpha(u_i)$ is a partition of U .

Definition 5 Given $IVDS = (U, A \cup \{d\})$ and $B \subseteq$

A , the information entropy of B is defined as follows:

$$\begin{aligned} H_{\text{SIM}}(B) &= - \sum_{i=1}^{|U|} p\left(S_B^\alpha(u_i)\right) \log p\left(S_B^\alpha(u_i)\right) \\ &= - \sum_{i=1}^{|U|} \frac{|S_B^\alpha(u_i)|}{|U|} \log \frac{|S_B^\alpha(u_i)|}{|U|}, \end{aligned} \quad (5)$$

where $|\cdot|$ denotes the number of elements in the set.

At the same time, the conditional entropy of B to d is defined as follows:

$$\begin{aligned} H_{\text{SIM}}(d \mid B) &= - \sum_{i=1}^{|U|} p\left(S_B^\alpha(u_i)\right) \\ &\quad \cdot \sum_{j=1}^{|U/d|} p\left(D_j \mid S_B^\alpha(u_i)\right) \log p\left(D_j \mid S_B^\alpha(u_i)\right) \\ &= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \log \frac{|S_B^\alpha(u_i) \cap D_j|}{|S_B^\alpha(u_i)|}. \end{aligned} \quad (6)$$

The joint entropy of B and d is defined as follows:

$$\begin{aligned} H_{\text{SIM}}(d \cup B) &= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} p\left(D_j \cap S_B^\alpha(u_i)\right) \log p\left(D_j \cap S_B^\alpha(u_i)\right) \\ &= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \log \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|}. \end{aligned} \quad (7)$$

Note that we define $H_{\text{SIM}}(d \mid B) = 0$ when $|S_B^\alpha(u_i) \cap D_j| = 0$.

Proposition 4 Let $IVDS = (U, A \cup \{d\})$ be an interval-valued decision system. For the attribute subset $B \subseteq A$, we have

1. $H_{\text{SIM}}(B) \geq 0$;
2. $H_{\text{SIM}}(d \cup B) = \max\{H_{\text{SIM}}(d), H_{\text{SIM}}(B)\}$;
3. $H_{\text{SIM}}(d \mid B) = 0$ and $H_{\text{SIM}}(d \cup B) = H_{\text{SIM}}(B)$ if and only if $S_B^\alpha(u_i) \subseteq D(u_i), \forall u_i \in U$;
4. $H_{\text{SIM}}(d \cup B) = H_{\text{SIM}}(d \mid B) + H_{\text{SIM}}(B)$.

Proof

$$\begin{aligned}
& H_{\text{SIM}}(d \mid B) + H_{\text{SIM}}(B) \\
&= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \log \frac{|S_B^\alpha(u_i) \cap D_j|}{|S_B^\alpha(u_i)|} \\
&\quad - \sum_{i=1}^{|U|} \frac{|S_B^\alpha(u_i)|}{|U|} \log \frac{|S_B^\alpha(u_i)|}{|U|} \\
&= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \left(\log \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \right. \\
&\quad \left. - \log \frac{|S_B^\alpha(u_i)|}{|U|} \right) \\
&\quad - \sum_{i=1}^{|U|} \frac{|S_B^\alpha(u_i)|}{|U|} \log \frac{|S_B^\alpha(u_i)|}{|U|} \\
&= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \log \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \\
&\quad + \sum_{i=1}^{|U|} \left(\sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} - \frac{|S_B^\alpha(u_i)|}{|U|} \right) \\
&\quad \cdot \log \frac{|S_B^\alpha(u_i)|}{|U|} \\
&= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \log \frac{|S_B^\alpha(u_i) \cap D_j|}{|U|} \\
&\quad + \sum_{i=1}^{|U|} \left(\frac{|S_B^\alpha(u_i)|}{|U|} - \frac{|S_B^\alpha(u_i)|}{|U|} \right) \log \frac{|S_B^\alpha(u_i)|}{|U|} \\
&= H_{\text{SIM}}(d \cup B).
\end{aligned}$$

Note that

$$H_{\text{SIM}}(d \cup B) \neq H_{\text{SIM}}(B \mid d) + H_{\text{SIM}}(d).$$

4 Attribute reduction framework for interval-valued information systems and interval-valued decision systems based on information entropies

One significant problem in rough set theory is searching for particular subsets of attributes which provide the same information for classification. It is also called attribute reduction.

4.1 Information theory view for attribute reduction of interval-valued information systems

Definition 6 Assume that $\text{IVIS} = (U, A)$ is an interval-valued information system. If an attribute subset $B \subseteq A$ and $\alpha \in [0, 1]$, attribute subset B is a

reduct of IVIS if and only if

1. $S_B^\alpha = S_A^\alpha$;
2. $\forall b \in B, S_{B-\{b\}}^\alpha \neq S_A^\alpha$.

Definition 7 Let $\text{IVDS} = (U, A \cup \{d\})$ be an interval-valued decision system. If attribute subset $B \subseteq A$ and $\alpha \in [0, 1]$, attribute set B is a relative reduct of IVDS if and only if

1. $\partial_B^\alpha = \partial_A^\alpha$;
2. $\forall b \in B, \partial_{B-\{b\}}^\alpha \neq \partial_A^\alpha$.

To construct our information theory based attribute reduction methods, we provide an information theory view for attribute reduction in interval-valued information systems.

Theorem 1 Assume that $\text{IVIS} = (U, A)$ is an interval-valued information system. Then $S_B^\alpha = S_A^\alpha$ and $H_{\text{SIM}}(B) = H_{\text{SIM}}(A)$ are equivalent.

Proof

$$\begin{aligned}
& H_{\text{SIM}}(B) \\
&= - \sum_{i=1}^{|U|} \frac{|S_B^\alpha(u_i)|}{|U|} \log \frac{|S_B^\alpha(u_i)|}{|U|} \\
&= - \sum_{i=1}^{|U|} \frac{|S_A^\alpha(u_i)|}{|U|} \log \frac{|S_A^\alpha(u_i)|}{|U|} = H_{\text{SIM}}(A).
\end{aligned}$$

Theorem 2 Let $\text{IVIS} = (U, A)$ be an interval-valued information system. If $C \subseteq A$ is redundant, then $H_{\text{SIM}}(A) - H_{\text{SIM}}(A - C) = 0$.

Proof If C is redundant in IVIS, then we know $S_{\{A-C\}}^\alpha = S_A^\alpha$ by Definition 6. It is easy to obtain $H_{\text{SIM}}(A - C) = H_{\text{SIM}}(A)$ by Theorem 1.

Theorem 3 Let $\text{IVIS} = (U, A)$ be an interval-valued information system. Then $C \subseteq A$ is indispensable if and only if $H_{\text{SIM}}(A) - H_{\text{SIM}}(A - C) > 0$.

Proof Suppose the attribute subset C is indispensable. Then we have $S_{\{A-C\}}^\alpha \neq S_A^\alpha$. By Theorem 1, we have $H_{\text{SIM}}(A - C) \neq H_{\text{SIM}}(A)$. Since $H_{\text{SIM}}(A - C) \leq H_{\text{SIM}}(A)$, it follows that $H_{\text{SIM}}(A) - H_{\text{SIM}}(A - C) > 0$.

Definition 8 Assume that $\text{IVIS} = (U, A)$ is an interval-valued information system. The significance of an attribute $a_i \notin B$ relative to B is as follows:

$$\text{Sig}(a_i, B) = H_{\text{SIM}}(B) - H_{\text{SIM}}(B - \{a_i\}). \quad (8)$$

Definition 9 If $\text{IVDS} = (U, A \cup \{d\})$ is an interval-valued decision system, the significance of

attribute a_i relative to B is defined as

$$\text{Sig}(a_i, B, d) = H_{\text{SIM}}(d \mid B - \{a_i\}) - H_{\text{SIM}}(d \mid B). \quad (9)$$

From the above theorems, we obtain the information theory view for attribute reduction in interval-valued information systems.

Theorem 4 Assume that $\text{IVIS} = (U, A)$ is an interval-valued information system. The attribute subset $B \subseteq A$ is a reduct of A if and only if

1. $H_{\text{SIM}}(B) = H_{\text{SIM}}(A)$;
2. $\forall b \in B, \text{Sig}(b, B, A) > 0$.

Theorem 5 Given $\text{IVDS} = (U, A \cup \{d\})$ and $B \subseteq A$, if IVDS is consistent, then $\partial_B^\alpha = \partial_A^\alpha$ and $H_{\text{SIM}}(d \mid B) = H_{\text{SIM}}(d \mid A)$ are equivalent.

Proof Assume $\partial_B^\alpha = \partial_A^\alpha$. Since $\text{IVDS} = \{U, A \cup \{d\}\}$ is consistent, we have $|\partial_A^\alpha(u_i)| = 1$ and $S_A^\alpha(u_i) \subseteq D(u_i)$ for all objects $u_i \in U$. Hence, $|\partial_B^\alpha(u_i)| = 1$ and $S_B^\alpha(u_i) \subseteq D(u_i)$ for all objects $u_i \in U$. By Proposition 4, we have $H(d \mid B) = 0$. Since $H(d \mid A) = 0$, we have $H(d \mid B) = H(d \mid A)$.

Suppose $H(d \mid B) = H(d \mid A)$. Since IVDS is consistent, we have $H(d \mid B) = 0$. Hence, $S_B^\alpha(u_i) \subseteq D(u_i)$ for all objects $u_i \in U$. Since $S_A^\alpha(u_i) \subseteq D(u_i)$, we have $\partial_B^\alpha = \partial_A^\alpha$. This completes the proof.

Theorem 6 Let $\text{IVDS} = (U, A \cup \{d\})$ be a consistent interval-valued decision system. Then $C \subseteq A$ is dispensable if and only if $H_{\text{SIM}}(d \mid A - C) = 0$.

Proof Suppose that $\text{IVDS} = (U, A \cup \{d\})$ is a consistent interval-valued decision system and that C is a dispensable attribute subset. Then we have $\partial_{A-C}^\alpha = \partial_A^\alpha$ by Definition 7. According to Theorem 5, we obtain $H_{\text{SIM}}(d \mid A - C) = H_{\text{SIM}}(d \mid A)$. Since IVDS is consistent, we have $H_{\text{SIM}}(d \mid A) = 0$. Therefore, it follows that $H_{\text{SIM}}(d \mid A - C) = 0$.

Theorem 7 Let $\text{IVDS} = (U, A \cup \{d\})$ be a consistent interval-valued decision system. Then $C \subseteq A$ is indispensable if and only if $H_{\text{SIM}}(d \mid A - C) > 0$.

Proof Assume that $\text{IVDS} = (U, A \cup \{d\})$ is a consistent interval-valued decision system and that the attribute subset C is indispensable. Then we have $\partial_{A-C}^\alpha \neq \partial_A^\alpha$. It is not difficult to obtain $H_{\text{SIM}}(d \mid A - C) \neq H_{\text{SIM}}(d \mid A)$. Then we have $H_{\text{SIM}}(d \mid A - C) \geq H_{\text{SIM}}(d \mid A)$ and $H_{\text{SIM}}(d \mid A) = 0$, according to Proposition 4. Consequently, we have $H_{\text{SIM}}(d \mid A - C) > 0$.

From the above theorems, we obtain the information theory view for attribute reduction in consistent interval-valued decision systems.

Theorem 8 Assume that $\text{IVDS} = (U, A \cup \{d\})$ is a consistent interval-valued decision system. The attribute subset $B \subseteq A$ is a reduct of IVDS if and only if

1. $H_{\text{SIM}}(d \mid B) = H_{\text{SIM}}(d \mid A)$;
2. $\forall b \in B, \text{Sig}(b, B, d) > 0$.

4.2 Attribute reduction algorithms for IVIS and IVDS

The attribute reduction algorithms for IVIS and IVDS are given in Algorithms 1 and 2, respectively.

Algorithm 1 Attribute reduction algorithm for an IVIS

```

1: Initialize the set of attribute reduct to the empty set,
   i.e.,  $B = \emptyset$ .
2: while  $H_{\text{SIM}}(B) \neq H_{\text{SIM}}(A)$  do
3:    $\hat{a}_i = \max_{a_i \in A-B} H_{\text{SIM}}(B + \{a_i\})$ 
4:    $B = B \cup \{\hat{a}_i\}$ 
5: end while

```

Algorithm 2 Attribute reduction algorithm for a consistent IVDS

```

1: Initialize the set of attribute reduct to the empty set,
   i.e.,  $B = \emptyset$ .
2: while  $H(d \mid B) \neq 0$  do
3:    $\hat{a}_i = \min_{a_i \in A-B} H(d \mid B + \{a_i\})$ 
4:    $B = B \cup \{\hat{a}_i\}$ 
5: end while

```

Now we provide an example to illustrate the proposed method. Assume that an IVDS is as listed in Table 1.

Table 1 An example with interval values for attribute reduction

U	a	b	c	d
u_1	[2.17, 2.86]	[2.45, 2.96]	[5.32, 7.230]	1
u_2	[3.37, 4.75]	[3.43, 4.85]	[7.24, 10.47]	2
u_3	[1.83, 2.70]	[1.78, 2.98]	[7.23, 10.27]	3
u_4	[1.35, 2.12]	[1.42, 2.09]	[2.59, 3.930]	4
u_5	[3.46, 5.35]	[3.37, 5.11]	[6.37, 10.28]	2
u_6	[2.22, 3.07]	[2.43, 3.32]	[4.37, 7.050]	1
u_7	[2.51, 4.04]	[2.52, 4.12]	[7.12, 11.26]	2
u_8	[1.00, 1.72]	[1.10, 1.82]	[3.58, 5.650]	4

Example 2 Suppose the similarity rate $\alpha = 0.8$ and $A = \{a, b, c\}$. We confirm that the IVDS shown in Table 1 is consistent when $\alpha = 0.8$. Let us

compute $S_A^{0.8}(u_1)$ in detail. By Definition 1, we have

$$\begin{aligned} v_a(u_1, u_1) &= 1.0000, & v_a(u_1, u_2) &= 0.6448, \\ v_a(u_1, u_3) &= 0.9425, & v_a(u_1, u_4) &= 0.8207, \\ v_a(u_1, u_5) &= 0.5655, & v_a(u_1, u_6) &= 0.9701, \\ v_a(u_1, u_7) &= 0.8253, & v_a(u_1, u_8) &= 0.7345. \end{aligned}$$

Then, we have

$$S_a^{0.8}(u_1) = \{u_1, u_3, u_4, u_6, u_7\}.$$

Similarly, we have

$$\begin{aligned} S_b^{0.8}(u_1) &= \{u_1, u_3, u_6, u_7\}, \\ S_c^{0.8}(u_1) &= \{u_1, u_6, u_8\}. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} S_A^{0.8}(u_1) &= S_a^{0.8}(u_1) \cap S_b^{0.8}(u_1) \cap S_c^{0.8}(u_1) \\ &= \{u_1, u_3, u_4, u_6, u_7\} \cap \{u_1, u_3, u_6, u_7\} \cap \{u_1, u_6, u_8\} \\ &= \{u_1, u_6\}. \end{aligned}$$

In the same way, we can compute $S_A^{0.8}(u_i)$ for all objects $u_i \in U$:

$$\begin{aligned} S_A^{0.8}(u_1) &= \{u_1, u_6\}, & S_A^{0.8}(u_5) &= \{u_2, u_5\}, \\ S_A^{0.8}(u_2) &= \{u_2, u_5\}, & S_A^{0.8}(u_6) &= \{u_1, u_6\}, \\ S_A^{0.8}(u_3) &= \{u_3, u_6\}, & S_A^{0.8}(u_7) &= \{u_7\}, \\ S_A^{0.8}(u_4) &= \{u_4, u_8\}, & S_A^{0.8}(u_8) &= \{u_4, u_8\}. \end{aligned}$$

By decision attribute d , we have the partition

$$U/d = \left\{ \{u_1, u_6\}, \{u_3\}, \{u_2, u_5, u_7\}, \{u_4, u_8\} \right\}.$$

The decision classes based on the decision attribute are also listed as follows:

$$\begin{aligned} D_1 &= \{u_1, u_6\}, & D_3 &= \{u_3\}, \\ D_2 &= \{u_2, u_5, u_7\}, & D_4 &= \{u_4, u_8\}. \end{aligned}$$

First, an empty set is assigned to the original attribute reduction set B , i.e., $B = \emptyset$, and the conditional entropy for each single attribute a , b , or c can be computed.

By definition, we have

$$\begin{aligned} H_{\text{SIM}}(d | a) &= - \sum_{i=1}^{|U|} p \left(S_a^\alpha(u_i) \right) \\ &\quad \cdot \sum_{j=1}^{|U/d|} p \left(D_j | S_a^\alpha(u_i) \right) \log p \left(D_j | S_a^\alpha(u_i) \right) \\ &= - \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|S_a^\alpha(u_i) \cap D_j|}{|U|} \log \frac{|S_a^\alpha(u_i) \cap D_j|}{|S_a^\alpha(u_i)|} \\ &= - \sum_{i=1}^{|U|} \left[\frac{|S_a^{0.8}(u_i) \cap D_1|}{|U|} \log \frac{|S_a^{0.8}(u_i) \cap D_1|}{|S_a^{0.8}(u_i)|} \right. \\ &\quad + \frac{|S_a^{0.8}(u_i) \cap D_2|}{|U|} \log \frac{|S_a^{0.8}(u_i) \cap D_2|}{|S_a^{0.8}(u_i)|} \\ &\quad + \frac{|S_a^{0.8}(u_i) \cap D_3|}{|U|} \log \frac{|S_a^{0.8}(u_i) \cap D_3|}{|S_a^{0.8}(u_i)|} \\ &\quad \left. + \frac{|S_a^{0.8}(u_i) \cap D_4|}{|U|} \log \frac{|S_a^{0.8}(u_i) \cap D_4|}{|S_a^{0.8}(u_i)|} \right] \\ &= 3.9512. \end{aligned}$$

Hence, we obtain

$$\begin{aligned} H_{\text{SIM}}(d | a) &= 3.9512, \\ H_{\text{SIM}}(d | b) &= 2.9387, \\ H_{\text{SIM}}(d | c) &= 2.8113. \end{aligned}$$

Then, we find $H_{\text{SIM}}(d | c)$ is the minimum, and attribute c is added to attribute reduction set B , i.e., $B = \{c\}$. The next step is to compute the conditional entropies of $c \cup a$ and $c \cup b$:

$$H_{\text{SIM}}(d | a \cup c) = 0, \quad H_{\text{SIM}}(d | b \cup c) = 0.$$

Since $H_{\text{SIM}}(d | a \cup c) = H_{\text{SIM}}(d | b \cup c)$, we choose the one explored first, i.e., attribute a . The algorithm terminates since the conditional entropy is equal to zero. Hence, we obtain the reduct $B = \{a, c\}$.

5 Experiments

To test the effectiveness of the proposed algorithm, experiments on three real-world datasets are performed. All values of conditional attributes in the datasets are interval values.

5.1 Fish dataset (or ecotoxicology dataset)

The Fish dataset has been introduced to test the effectiveness of attribute reduction for symbolic

interval data (Hedjazi *et al.*, 2011). This dataset is composed of observations for abnormal levels of mercury contamination in some Amerindian areas, taken from several studies in French Guyana by researchers from the LEESA Laboratory.

There are 13 interval conditional attributes, which are length, weight, muscle, intestine, stomach, gills, liver, kidneys, liver/muscle, kidneys/muscle, gills/muscle, intestine/muscle, and stomach/muscle. Similarly, they are labeled as a_1 – a_{13} . What is more, a reference classification with respect to the fish diet is taken as the decision attribute, including carnivorous, detritivorous, omnivorous, and herbivorous.

The conditional entropy for each conditional attribute is illustrated in Fig. 1. We can find that the conditional entropy of attribute a_9 , which represents the liver/muscle, is the minimum. This means liver/muscle is more significant than the other attributes.

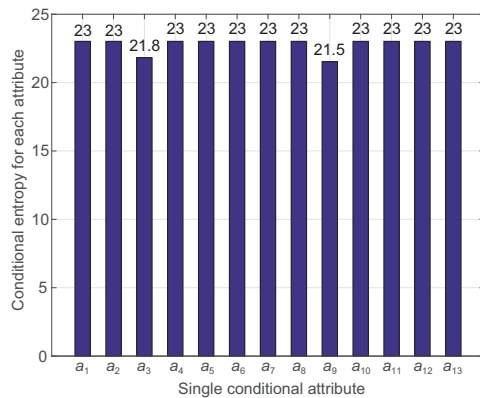


Fig. 1 Conditional entropy for the single attribute of the Fish dataset

5.2 Face Recognition dataset

The Face Recognition dataset focuses on face recognition. Each interval value represents the measurement of each local feature in a face image. For each face image, the localization of the salient features such as nose, mouth, and eyes is obtained by using morphological operators. A distance is measured between specific points delimiting each boundary and several distances are described as interval values.

The dataset contains 9 men with 3 sequences for each, giving a total of 27 observations (Dai *et al.*, 2013b). The decision attribute identifies which per-

son it is. There are 6 conditional attributes, including the length spanned by the eyes, the length between the eyes, the length from the outer right eye to the upper middle lip at the point between the nose and mouth, and so on. We can represent the six conditional attributes as a_1 – a_6 .

By using our reduction algorithm, first the set of attribute reductions, B , is initialized to the empty set, and we can compute the conditional entropy for each single conditional attribute shown in Fig. 2. Then, we can find that the conditional entropy of attribute a_3 (which represents the length from the outer right eye to the upper middle lip at the point between the nose and the mouth) is the minimum. This means the length from the outer right eye to the upper middle lip at the point between the nose and the mouth is more significant than the other attributes.

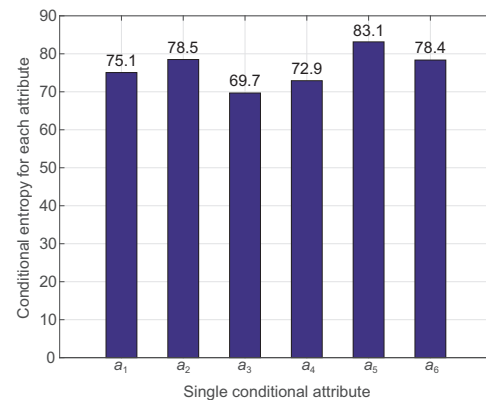


Fig. 2 Conditional entropy for the single attribute of the Face Recognition dataset

5.3 Car dataset

The Car dataset contains 33 car models described by 7 interval variables, 2 categorical multi-valued variables, and 1 nominal variable. This dataset has been used in research on clustering for interval values (Hedjazi *et al.*, 2011; Dai *et al.*, 2013b).

In this study, we take only the 7 interval conditional attributes into consideration, which are price, engine capacity, top speed, step, length, width, and height, denoted as a_1 – a_7 . The nominal variable ‘car category’ represents the decision attribute, which has been used as a priori classification.

In Fig. 3, we list the conditional entropy of each attribute a_1 – a_7 to decision attribute d . It is easy

to find that $H_{\text{SIM}}(d | a_1)$ is the minimum, and we can conclude that the attribute price provides more information for classification.

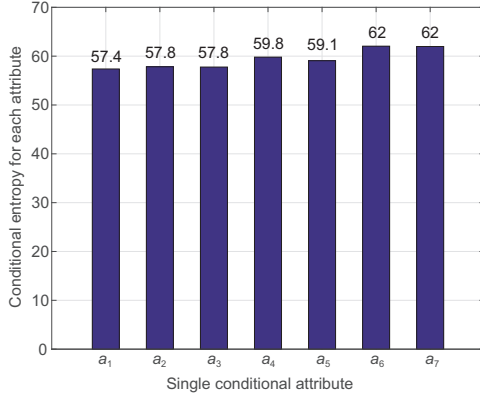


Fig. 3 Conditional entropy for the single attribute of the Car dataset

5.4 Comparison of performance

To verify the effectiveness of the proposed approach, attribute selection experiments based on the proposed uncertainty measurements are conducted on the above datasets.

Besides the similarity measure constructed in Definition 1, we consider two other similarity measures for comparison. One is constructed by Dai et al. (2012; 2013b) based on possible degree. The similarity is called possible degree similarity. Let $A = [a^-, a^+]$ and $B = [b^-, b^+]$ be two interval values. The possible degree similarity between the two interval values is defined as

$$S_{AB} = 1 - \left| P_{(A \geq B)} - P_{(B \geq A)} \right|, \quad (10)$$

where

$$P_{(A \geq B)} = \min \left\{ 1, \max \left\{ \frac{a^+ - b^-}{(a^+ - a^-) + (b^+ - b^-)}, 0 \right\} \right\}. \quad (11)$$

$P_{(A \geq B)}$ and $P_{(B \geq A)}$ are the possible degree of A relative to B and the possible degree of B relative to A , respectively.

The other similarity measure can be adapted from that used by Dai and Tian (2013) to handle set-valued data. For interval values $A = [a^-, a^+]$ and $B = [b^-, b^+]$, the intersection-union similarity is

defined as

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|[a^-, a^+] \cap [b^-, b^+]|}{|[a^-, a^+] \cup [b^-, b^+]|}. \quad (12)$$

It is easy to prove that the similarity has the three properties in Proposition 1.

Very few classifiers can be used to address interval data; therefore, to compare the performance of classification based on the selected attributes, Dai et al. (2013b) extended the classical k -nearest neighbor (KNN) classifier and probabilistic neural network (PNN) classifier to handle interval-valued data by redefining the distance between two objects:

Definition 10 (Dai et al., 2013b) Suppose that X and Y are two objects in interval-valued information systems and that u_i^κ and u_j^κ are two interval values at the κ th attribute. The distance between X and Y is defined as follows:

$$\text{Dis}(X, Y) = \sqrt{\sum_{\kappa=1}^m \left(P_{(u_i^\kappa \geq u_j^\kappa)} - P_{(u_j^\kappa \geq u_i^\kappa)} \right)^2}, \quad (13)$$

where m is the number of conditional attributes and $P_{(u_i^\kappa \geq u_j^\kappa)}$ is the possible degree between two interval values.

Due to the limited number of objects in the dataset, we use a leave-one-out cross-validation approach to evaluate the classification performances.

In the experiments, the relative bound difference similarity, possible degree similarity, and intersection-union similarity are used in the proposed attribute reduction framework, denoted as RBD, PD, and IU, respectively. We also compare these three methods with the attribute selection method based on uncertainty measurement for interval-valued information systems constructed by Dai et al. (2013b), called uncertainty measurement attribute reduction (UMAR).

The results are shown in Tables 2–4. The parameter α represents the similarity threshold (Section 2.2), used to construct similarity relations and similarity classes. Tables 2–4 show the accuracy rates on the Fish, Face Recognition, and Car datasets, respectively, obtained by the extended KNN classifier and the extended PNN classifier. From the results, we find that RBD obviously outperforms the existing method UMAR. Moreover, PD is a little better than UMAR. IU performs as well as UMAR. The results indicate that the proposed attribute reduction method is feasible and effective.

Table 2 Performance on the Fish dataset by KNN and PNN

Threshold	RBD		IU		PD		UMAR	
α	KNN	PNN	KNN	PNN	KNN	PNN	KNN	PNN
Full attributes	0.5833	0.6667	0.5833	0.6667	0.5833	0.6667	0.5833	0.6667
0.10	0.5833	0.6667	0.5000	0.5000	0.5000	0.5000	0.5833	0.5833
0.15	0.5833	0.6667	0.5000	0.5000	0.5000	0.5000	0.5833	0.5833
0.20	0.5833	0.6667	0.6667	0.6667	0.5000	0.5000	0.5833	0.5833
0.25	0.5833	0.6667	0.6667	0.6667	0.6667	0.6667	0.5000	0.5000
0.30	0.5833	0.6667	0.5000	0.5833	0.6667	0.6667	0.5000	0.5000
0.35	0.5833	0.6667	0.5000	0.5833	0.6667	0.6667	0.5000	0.5000
0.40	0.3333	0.5833	0.5000	0.5000	0.6667	0.6667	0.5000	0.5000
0.45	0.5833	0.7500	0.5000	0.5000	0.6667	0.6667	0.5000	0.5000
0.50	0.5833	0.8333	0.3333	0.4167	0.5000	0.5833	0.5000	0.5000
0.55	0.5833	0.5833	0.3333	0.4167	0.5000	0.5833	0.5000	0.5000
0.60	0.5000	0.5000	0.3333	0.4167	0.5000	0.5000	0.3333	0.4167
0.65	0.5833	0.5833	0.1667	0.3333	0.3333	0.4167	0.3333	0.4167
0.70	0.5000	0.5833	0.1667	0.3333	0.3333	0.4167	0.3333	0.4167
0.75	0.5000	0.5000	0.3333	0.4167	0.3333	0.4167	0.3333	0.4167

Table 3 Performance on the Face Recognition dataset by KNN and PNN

Threshold	RBD		IU		PD		UMAR	
α	KNN	PNN	KNN	PNN	KNN	PNN	KNN	PNN
Full attributes	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.10	0.4074	0.7407	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.15	0.9629	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9259	1.0000
0.40	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9259	1.0000
0.45	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9259	1.0000
0.50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9259	1.0000
0.55	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9259	1.0000
0.60	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.65	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.70	1.0000	1.0000	0.7407	0.9259	1.0000	1.0000	1.0000	1.0000
0.75	1.0000	1.0000	0.7407	0.9259	1.0000	1.0000	0.9259	1.0000

Table 4 Performance on the Car dataset by KNN and PNN

Threshold	RBD		IU		PD		UMAR	
α	KNN	PNN	KNN	PNN	KNN	PNN	KNN	PNN
Full attributes	0.6364	0.6970	0.6364	0.6970	0.6364	0.6970	0.6364	0.6970
0.10	0.8182	0.8182	0.6061	0.6364	0.6364	0.6364	0.6970	0.7273
0.15	0.7576	0.7576	0.6061	0.6364	0.6364	0.6364	0.6970	0.7273
0.20	0.7576	0.7576	0.7273	0.6970	0.6364	0.6364	0.6970	0.7273
0.25	0.8182	0.8182	0.7273	0.6970	0.6364	0.6364	0.6970	0.7273
0.30	0.8182	0.8182	0.6061	0.6364	0.6364	0.6364	0.6970	0.7273
0.35	0.6364	0.6970	0.7273	0.6970	0.6364	0.6364	0.6970	0.7273
0.40	0.6364	0.6970	0.7273	0.6970	0.6364	0.6364	0.6970	0.7273
0.45	0.7576	0.6970	0.7273	0.6970	0.6364	0.6364	0.6970	0.7273
0.50	0.7576	0.6970	0.6061	0.6667	0.6364	0.6364	0.6970	0.7273
0.55	0.7576	0.6970	0.6061	0.6667	0.6364	0.6364	0.6970	0.7273
0.60	0.7576	0.6970	0.5455	0.6061	0.6364	0.6364	0.5758	0.5455
0.65	0.6364	0.6970	0.6364	0.6364	0.6364	0.6364	0.4848	0.6061
0.70	0.6667	0.7879	0.6364	0.6364	0.6364	0.6364	0.4848	0.6061
0.75	0.6364	0.6970	0.6364	0.6364	0.6364	0.6364	0.4848	0.6061

Applying different similarity measures to the proposed algorithm leads to different results. Among RBD, IU, and PD, RBD outperforms IU and PD. In most cases, RBD obtains the best results. PD performs a little better than IU. The reason may lie in the differences of the abilities of describing the similarity between two interval values among these three measures.

6 Conclusions

The classical rough set model is not appropriate for handling interval-valued data. In this paper, we present a new framework for attribute reduction in interval-valued information systems from the viewpoint of information theory. Some information theory concepts, including entropy, conditional entropy, and joint entropy, are defined in interval-valued information systems. Based on these concepts, we provide an information theory view for attribute reduction in interval-valued information systems. Consequently, attribute reduction algorithms are proposed. To test the proposed algorithms, experiments on three datasets are conducted. Experiments show that the proposed framework is effective for attribute reduction in interval-valued information systems.

In the future, we plan to investigate other similarity measures and introduce them into our information theory framework of attribute reduction for interval-valued information systems.

References

- Billard, L., Douzal-Chouakria, A., Diday, E., 2008. Symbolic Principal Component for Interval-Valued Observations. Available from <https://hal.archives-ouvertes.fr/hal-00361053>.
- Bustince, H., Barrenechea, E., Pagola, M., 2006. Restricted equivalence functions. *Fuzzy Sets Syst.*, **157**(17):2333-2346. <http://dx.doi.org/10.1016/j.fss.2006.03.018>
- Dai, J.H., 2008. Rough 3-valued algebras. *Inform. Sci.*, **178**(8):1986-1996. <http://dx.doi.org/10.1016/j.ins.2007.11.011>
- Dai, J.H., Tian, H.W., 2013. Fuzzy rough set model for set-valued data. *Fuzzy Sets Syst.*, **229**:54-68. <http://dx.doi.org/10.1016/j.fss.2013.03.005>
- Dai, J.H., Xu, Q., 2012. Approximations and uncertainty measures in incomplete information systems. *Inform. Sci.*, **198**:62-80. <http://dx.doi.org/10.1016/j.ins.2012.02.032>
- Dai, J.H., Xu, Q., 2013. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Appl. Soft Comput.*, **13**(1):211-221. <http://dx.doi.org/10.1016/j.asoc.2012.07.029>
- Dai, J.H., Wang, W.T., Xu, Q., et al., 2012. Uncertainty measurement for interval-valued decision systems based on extended conditional entropy. *Knowl.-Based Syst.*, **27**:443-450. <http://dx.doi.org/10.1016/j.knsys.2011.10.013>
- Dai, J.H., Tian, H.W., Wang, W.T., et al., 2013a. Decision rule mining using classification consistency rate. *Knowl.-Based Syst.*, **43**:95-102. <http://dx.doi.org/10.1016/j.knsys.2013.01.010>
- Dai, J.H., Wang, W.T., Mi, J.S., 2013b. Uncertainty measurement for interval-valued information systems. *Inform. Sci.*, **251**:63-78. <http://dx.doi.org/10.1016/j.ins.2013.06.047>
- Dai, J.H., Wang, W.T., Xu, Q., 2013c. An uncertainty measure for incomplete decision tables and its applications. *IEEE Trans. Cybern.*, **43**(4):1277-1289. <http://dx.doi.org/10.1109/TSMCB.2012.2228480>
- Galar, M., Fernandez, J., Beliakov, G., et al., 2011. Interval-valued fuzzy sets applied to stereo matching of color images. *IEEE Trans. Image Process.*, **20**(7):1949-1961. <http://dx.doi.org/10.1109/TIP.2011.2107525>
- Hedjazi, L., Aguilar-Martin, J., Le Lann, M.V., 2011. Similarity-margin based feature selection for symbolic interval data. *Patt. Recogn. Lett.*, **32**(4):578-585. <http://dx.doi.org/10.1016/j.patrec.2010.11.018>
- Hu, Y.C., 2015. Flow-based tolerance rough sets for pattern classification. *Appl. Soft Comput.*, **27**:322-331. <http://dx.doi.org/10.1016/j.asoc.2014.11.021>
- Kryszkiewicz, M., 1998. Rough set approach to incomplete information systems. *Inform. Sci.*, **112**(1-4):39-49. [http://dx.doi.org/10.1016/S0020-0255\(98\)10019-1](http://dx.doi.org/10.1016/S0020-0255(98)10019-1)
- Leung, Y., Fischer, M.M., Wu, W.Z., et al., 2008. A rough set approach for the discovery of classification rules in interval-valued information systems. *Int. J. Approx. Reason.*, **47**(2):233-246. <http://dx.doi.org/10.1016/j.ijar.2007.05.001>
- Lin, S.H., Huang, C.C., Che, Z.X., 2015. Rule induction for hierarchical attributes using a rough set for the selection of a green fleet. *Appl. Soft Comput.*, **37**:456-466. <http://dx.doi.org/10.1016/j.asoc.2015.08.016>
- Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht & Boston. <http://dx.doi.org/10.1007/978-94-011-3534-4>
- Qian, Y.H., Liang, J.Y., Dang, C.Y., 2008. Interval ordered information systems. *Comput. Math. Appl.*, **56**(8):1994-2009. <http://dx.doi.org/10.1016/j.camwa.2008.04.021>
- Yang, X.B., Yu, D.J., Yang, J.Y., et al., 2009. Dominance-based rough set approach to incomplete interval-valued information system. *Data Knowl. Eng.*, **68**(11):1331-1347. <http://dx.doi.org/10.1016/j.datak.2009.07.007>
- Zhang, C.Y., Fu, H.Y., 2006. Similarity measures on three kinds of fuzzy sets. *Patt. Recogn. Lett.*, **27**(12):1307-1317. <http://dx.doi.org/10.1016/j.patrec.2005.11.020>
- Zhang, X.H., Dai, J.H., Yu, Y.C., 2015. On the union and intersection operations of rough sets based on various approximation spaces. *Inform. Sci.*, **292**:214-229. <http://dx.doi.org/10.1016/j.ins.2014.09.007>
- Zhang, X.H., Miao, D.Q., Liu, C.H., et al., 2016. Constructive methods of rough approximation operators and multigranulation rough sets. *Knowl.-Based Syst.*, **91**:114-125. <http://dx.doi.org/10.1016/j.knsys.2015.09.036>