Entropy measures and granularity measures for set-valued information systems

Jianhua Dai *, Haowei Tian

College of Computer Science, Zhejiang University, Hangzhou 310027, China

Article info
Article history:
Received 12 October 2011
Received in revised form 30 December 2012
Accepted 25 March 2013
Available online 1 April 2013

Keywords:
Set-valued information system
Entropy
Granularity
Measure
Uncertainty

Abstract
Set-valued information systems are generalized models of single-valued information systems. In this paper, we propose two new relations for set-valued information systems. Based on these two relations, the concepts of knowledge information entropy, knowledge rough entropy, knowledge granulation and knowledge granularity measure are defined in set-valued information systems, and some properties are investigated. Moreover, relationship between knowledge information entropy and knowledge granulation, knowledge rough entropy and knowledge granularity measure are studied. It is also shown that knowledge information entropy and knowledge granularity measure can be used to evaluate the certainty degree of knowledge in set-valued information systems, and knowledge rough entropy and knowledge granulation can be used to evaluate the uncertainty degree of knowledge in set-valued information systems. These results may supply a further understanding the essence of uncertainty and granularity in set-valued information systems.

1. Introduction

Rough set theory is a mathematical tool for dealing with uncertain or imprecise information [35]. It has attracted the attention of many researchers who have studied its theories and its applications during the last decades [1,4–7,10,11,13,17–21,23,29,32,34,44,45,47,48,50,54,56,59,61].

Classical rough set philosophy is based on an assumption that every object in the universe of discourse is associated with some information (knowledge), expressed by means of some attributes used for object description [35]. Objects characterized by the same information are indiscernible. The indiscernibility relation generated in this way forms the mathematical basis for the theory of rough sets. In many practical issues, it may happen that some of the attribute values for an object are set-valued, which are always used to characterize uncertain information and missing information in information systems. Set-valued information systems can be viewed as generalized models of single-valued information systems [15]. Moreover, set-valued information systems can be used to handle incomplete information systems, in which all missing values can be represented by the set of all possible values of each attribute [38,51,55].

Proposed by Shannon [42] to evaluate uncertainty of a system, entropy has been applied in diverse fields as a very useful mechanism for characterizing information contents in various modes. The extension of entropy and its variants were adapted for rough sets in [2,8,9,12,14,24,36,39,46]. For example, Duentsch and Gediga defined the information entropy and three kinds of conditional entropies in rough sets for predicting a decision attribute [14]. Beaubouef et al. [2] proposed a method measuring uncertainty of rough sets and rough relation databases based on rough entropy. Wierman [46] presented the measures of uncertainty and granularity in rough set theory, along with an axiomatic derivation. Liang et al.
proposed a new method for evaluating both uncertainty and fuzziness. Qian and Liang proposed a combination entropy for evaluating uncertainty of a knowledge from an information system. All these studies were dedicated to evaluating uncertainty of a set in terms of the partition ability of knowledge. As a powerful mechanism, granulation was introduced by Zadeh. As a recently renewed research topic, granular computing concerns problem solving and information processing at multiple levels of granularity. Rough set model can be viewed as an example of partition based granular computing model. From the viewpoint of granulation, Yao defined a granularity measure in. Several measures on knowledge in an information system were proposed and studied by Liang et al. These measures include granularity measure, information entropy, rough entropy, and knowledge granulation. Qian et al. studied knowledge granulation in a knowledge base, and fuzzy information granularity in a binary granular structure. Xu et al. introduced concepts of knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems. Zhu et al. developed a pair of information-theoretic entropy and co-entropy functions associated to partitions and approximations. For more details on this topic, one can refer to a systematic survey in. So far, however, uncertainty measurement in set-valued information systems has not been reported. In this paper, we aim to address uncertainty measurement issue in set-valued information systems. This paper introduces knowledge entropy, knowledge granulation, and roughness measure into set-valued information systems, and investigates some properties of them. It is shown that these proposed measures provide approaches to evaluate the discernibility ability of different knowledge in set-valued information systems.

The remainder of the paper is organized as follows. In Section 2, we briefly review some information-theoretic measures and granularity measures for rough sets in the literature. We propose two extended similarity relations for set-valued information systems in Section 3. Entropy measures including information entropy and rough entropy are introduced into set-valued information systems in Section 4. In Section 5, granularity measures including knowledge granulation and granularity measure are introduced into set-valued information systems. In Section 6, relationships between entropy measures and granularity measures are investigated. Section 7 concludes the paper.

2. Preliminaries

Shannon entropy has been widely used to measure the structuredness of attributes in databases and the nonspecificity of a finite set. There are several information-theoretic measures of uncertainty and granularity for rough sets, which are based upon the notion of entropy introduced by Shannon.

**Definition 2.1.** Let be an approximation space, where the partition consists of blocks , . The Shannon entropy of the partition is defined by

\[
H(\pi) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} \log_2 \frac{|A_i|}{|U|}
\]

The entropy reaches the maximum value for the finest partition consisting of singleton subsets of , and it reaches the minimum value 0 for the coarsest partition . In general, for two partitions with , we have . That is, the value of the entropy correctly reflects the order of partitions with respect to their granularity.

Duntch and Gediga and Miao et al. used Shannon entropy of a partition as a measure of its roughness or granularity. Wierman explicitly called Shannon entropy a granularity measure in. In this paper, we call granularity measure of partition , denoted as . Hence, we have

\[
G(\pi) = H(\pi) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} \log_2 \frac{|A_i|}{|U|}
\]

The Shannon entropy in Eq. (1) can be re-expressed as:

\[
H(\pi) = \log_2 |U| - \sum_{i=1}^{k} \frac{|A_i|}{|U|} \log_2 |A_i|
\]

Notice that the Hartley measure of uncertainty for a finite set is

\[
H(A_i) = \log_2 |A_i|
\]

It measures the amount of uncertainty associated with a finite set of possible alternatives, the nonspecificity inherent in the set.

The first term in Eq. (3) is exactly the Hartley measure of , which is a constant independent of any partition. The second term of the equation is basically an expectation of granularity with respect to all blocks in the partition. This quantity was first explicitly used by Yao to measure the granularity of a partition in. It should also be noticed that Beaubouef et al. implicitly treated it as a measure of granularity of a partition when defining the uncertainty of a rough set approximation. Liang and Shi called it the rough entropy of . Bianucci et al. called it co-entropy.
Definition 2.2. [3,25,53] Let \( \langle U, \pi \rangle \) be an approximation space, where the partition \( \pi \) consists of blocks \( A_i \), \( 1 \leq i \leq k \). The rough entropy \( E_r(\pi) \) of the partition \( \pi \) is defined by

\[
E_r(\pi) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} \log_2 \frac{|A_i|}{|U|} = -\sum_{i=1}^{k} \frac{|A_i|}{|U|} \log_2 \frac{1}{|A_i|}
\]

(5)

Besides the measures discussed above, there are some interaction-based measures which hinge on counting the number of interacting pairs of elements of a universal set under a partition [57].

Definition 2.3. [24] Let \( \langle U, \pi \rangle \) be an approximation space, where the partition \( \pi \) consists of blocks \( A_i \), \( 1 \leq i \leq k \). The information entropy \( E_i(\pi) \) of the partition \( \pi \) is defined by

\[
E_i(\pi) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} \log_2 \left( \frac{|A_i|}{|U|} \right) = \sum_{i=1}^{k} \frac{|A_i|}{|U|} \left( 1 - \frac{|A_i|}{|U|} \right)
\]

(6)

where \( A_i^c \) is the complement of \( A_i \), i.e. \( A_i^c = U - A_i \); \( \frac{1}{|U|} \) represents the probability of block \( A_i \) within the universe \( U \); \( \frac{|A_i|}{|U|} \) denotes the probability of the complement of \( A_i \) within the universe \( U \).

As pointed out in [24], the information entropy can measure both uncertainty and fuzziness in rough set theory.

Definition 2.4. [25,30] Let \( \langle U, \pi \rangle \) be an approximation space, where the partition \( \pi \) consists of blocks \( A_i \), \( 1 \leq i \leq k \). The information entropy \( GK(\pi) \) of the partition \( \pi \) is defined by

\[
GK(\pi) = \frac{1}{|U|^2} \sum_{i=1}^{k} |A_i|^2 = \frac{1}{|U|} \sum_{i=1}^{k} \frac{|A_i|^2}{|U|}
\]

(7)

The value \( |A_i \times A_j| \) is the number of interacting pairs in the block \( A_i \), the summation is the total number of interacting pairs induced by a partition, and \( |U \times U| \) is the number of interacting pairs induced by the coarsest partition corresponding to the equivalence relation \( U \times U \). The measure \( GK \) may be interpreted as a normalized cardinality of an equivalence relation. The minimum value \( 1/|U| \) is obtained from the finest partition \( \{\{x\}: x \in U\} \) and the maximum value 1 from the coarsest partition \( \{U\} \).

3. Set-valued information system

An information system is a quadruple \( IS = \langle U,A,V,f \rangle \), where the universe \( U \) is a non-empty finite set of objects, \( A \) is a non-empty finite set of attributes, \( V \) is the union of attribute domains, i.e. \( V = \bigcup_{a \in A} V_a \) and \( V_a \) is the set of all possible values for attribute \( a \in A \); \( f: U \times A \rightarrow V \) is a function which assigns particular values from domains of attributes to objects such as \( \forall a \in A, x \in U, f(a,x) \in V_a \). \( f(a,x) \) is the value of attribute \( a \) for object \( x \).

Yao et al. first explicitly proposed the notion of set-valued information system in [55], called “set-based information system” in their paper.

Definition 3.1. [51,55] A set-valued information system is defined to be quadruple

\( \langle U,A,V,f \rangle \)

where \( U \) is a non-empty finite set of objects, \( A \) is a non-empty finite set of attributes, \( V \) is the union of attribute domains, i.e. \( V = \bigcup_{a \in A} V_a \) and \( V_a \) is the set of all possible values for attribute \( a \in A \); \( f: U \times A \rightarrow 2^V \) is a function which assigns particular values from domains of attributes to objects such as \( \forall a \in A, x \in U, f(a,x) \in 2^V \). \( f(a,x) \) is the value of attribute \( a \) for object \( x \).

As Yao pointed out in [51], a set-valued information system is the same as a standard information system except that the information function is a set-valued function. Table 1 shows a set-valued information system.

The semantics of set-valued information systems have been studied by different approaches [27,28,33,43,51], which, actually, fall into the following two types [15]:

Type (a): For \( x \in U \) and \( a \in A \), \( a(x) \) is interpreted disjunctively. For example: If \( a \) is the attribute “majoring in”, then \( a(x) = \{\text{Arts}, \text{Biological Science}, \text{Computer Science}, \text{Economics}, \text{Mathematics}, \text{Physics}\} \) can be interpreted as: student \( x \) majors

<table>
<thead>
<tr>
<th>( U )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>(1,2)</td>
<td>(0,1,2)</td>
<td>(0,1)</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>(1)</td>
<td>(1)</td>
<td>(0,2)</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>(0,3)</td>
<td>(2)</td>
<td>(1,3)</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>(3)</td>
<td>(1,2)</td>
<td>(1,3)</td>
</tr>
</tbody>
</table>
in Arts, Biological Science, Computer Science, Economics, Mathematics or Physics, and x majors in only one of them. Incomplete information systems with some unknown attribute values or partial known attribute values are such type of set-valued information systems [27,28,33,43,51]. In an incomplete information system, we only know the unavailable value is in a set. Although an object must take exactly one value from its value domain, the available information may be insufficient for us to determine which value is the actual one. Instead, a set of values is used for the unknown value [51].

Type (b): For \( x \in U \) and \( a \in C \), \( \{x\} \) is interpreted conjunctively. For example: If \( a \) is the attribute "research interest", then \( a(x) = \{ \{\text{Data Mining}, \text{Evolutionary Computation}, \text{Machine Learning}, \text{Fuzzy Sets}, \text{Rough Sets}\}\} \) can be interpreted as: researcher \( x \) is interested in Data Mining, Evolutionary Computation, Machine Learning, Fuzzy Sets, Rough Sets. In other words, the research interests of \( x \) include Data Mining, Evolutionary Computation, Machine Learning, Fuzzy Sets, Rough Sets.

In this study, we mainly focus on semantic interpretation of type (b).

Yao defined a tolerance relation in set-valued information systems as follows [51]:

**Definition 3.2.** In a set-valued information system \( \langle U, A, V, f \rangle \), \( \forall b \in A \), a tolerance relation is defined as:

\[
T_b = \{ (x, y) | b(x) \cap b(y) \neq \emptyset \}
\]

(8)

For \( B \subseteq A \), a tolerance relation is defined as:

\[
T_B = \{ (x, y) | \forall b \in B, b(x) \cap b(y) \neq \emptyset \} \subseteq \bigcap_{b \in B} T_b
\]

(9)

When \( (x, y) \in T_B \), \( x \) and \( y \) are called indiscernible about \( B \) or \( x \) tolerant with \( y \) about \( B \).

It should be noticed that Guan et al. adopted Yao’s definition in [15]. Their definition shows a very weak condition for two objects to have a tolerance relation. In some cases, this definition is not reasonable, especially when the tolerance degree is not considered. For example, for an attribute \( b \), let \( b(x) = \{a_1, a_2, a_3, a_4, a_5, a_6\} \), and \( b(y) = \{a_0, a_2, a_5, a_9, a_{10}, a_{11}\} \). Then, by Definition 3.2, both \( (x, y) \) and \( (x, z) \) belong to \( T_b \). In other words, \( x \) and \( y \) are indiscernible, and at the same time \( x \) and \( z \) are indiscernible. However, it is obvious that it is more difficult to discern \( x \) from \( z \) than discern \( x \) from \( y \). This kind of information is not contained in Definition 3.2. In light of this, we suggest two extended relations for set-valued information systems.

**Definition 3.3.** In a set-valued information system \( \langle U, A, V, f \rangle \), for \( b \in A \) and \( \alpha \in [0, 1] \), a binary relation can be defined as:

\[
Q^\alpha_b = \{ (x, y) | \frac{|b(x) \cap b(y)|}{|b(x)|} \geq \alpha \}
\]

(10)

For \( B \subseteq A \), \( B = \{b_1, b_2, \ldots, b_m\} \) and \( \alpha = \{ \alpha_{b_1}, \alpha_{b_2}, \ldots, \alpha_{b_m} \} \), \( \forall i, \alpha_{b_i} \in [0, 1] \), a relation can be defined as:

\[
Q^\alpha_B = \{ (x, y) | \forall b_i \in B, \frac{|b_i(x) \cap b_i(y)|}{|b_i(x)|} \geq \alpha_{b_i} \} = \bigcap_{b_i \in B} Q^\alpha_{b_i}
\]

(11)

When \( (x, y) \in Q^\alpha_B \), \( x \) and \( y \) are called indiscernible with respect to \( B \).

The parameter \( \alpha \) is the degree of precision we required. The bigger the value of \( \alpha \) is, the finer the defined indiscernible relation \( Q \) is. \( \alpha \) can be used to adjust the precision degree of the relation \( Q \) between two set-valued objects.

**Proposition 3.1.** In a set-valued information system \( \langle U, A, V, f \rangle \), for \( b \in A, B \subseteq A \) and \( \alpha \in [0, 1] \), \( Q^\alpha_b \) and \( Q^\alpha_B \) are reflexive relations on \( U \).

**Proof.** For any \( x \in U \), we know \( \frac{|b(x) \cap b(x)|}{|b(x)|} = 1 \geq \alpha \). It follows that \( (x, x) \in Q^\alpha_b \). Hence, \( Q^\alpha_b \) is a reflexive relation. Similarly, we can prove that \( Q^\alpha_B \) is a reflexive relation too. \( \square \)

Let \( \langle U, A, V, f \rangle \) be a set-valued information system, \( x \in U, B \subseteq A \), then the indiscernible class of object \( x \) with respect to \( B \) under \( Q^\alpha_b \) is defined as:

\[
S_{Q^\alpha_b}(u) = \{ v \in U | (u, v) \in Q^\alpha_b \}
\]

(12)

There are two special cases of \( S_{Q^\alpha_b} \), i.e. the discrete case and the indiscrete case. The discrete case is defined as

\[
\overline{S}_{Q^\alpha_b} : \overline{S}_{Q^\alpha_b}(u) = \{ u \}, \forall u \in U
\]

(13)

The indiscrete case is defined as

\[
\overline{S}_{Q^\alpha_b} : \overline{S}_{Q^\alpha_b}(u) = U, \forall u \in U
\]

(14)

**Definition 3.4.** In a set-valued information system \( \langle U, A, V, f \rangle \), for \( b \in A \) and \( \alpha \in [0, 1] \), a binary relation can be defined as:
\[ T^2_b = \left\{ (x,y) \left| \frac{|b(x) \cap b(y)|}{|b(x) \cup b(y)|} \geq \alpha \right. \right\} \]  

(15)

For \( B \subseteq A, b = \{b_1, b_2, \ldots, b_m\} \) and \( \alpha = (\alpha_{b_1}, \alpha_{b_2}, \ldots, \alpha_{b_m}) \), \( \forall i, \alpha_{b_i} \in [0, 1] \), a relation can defined as:

\[ T^2_b = \left\{ (x,y) \left| \forall b_i \in B, \frac{|b_i(x) \cap b_i(y)|}{|b_i(x) \cup b_i(y)|} \geq \alpha_{b_i} \right. \right\} = \bigcap_{b_i \in B} T^2_{b_i} \]  

(16)

When \((x, y) \in Q_b, x \) and \( y \) are called indiscernible with respect to \( B \).

The parameter \( \alpha \) is the degree of precision we required. The bigger the value of \( \alpha \) is, the finer the defined indiscernible relation \( T \) is. \( \alpha \) can be used to adjust the precision degree of the relation \( T \) between two set-valued objects.

**Proposition 3.2.** In a set-valued information system \( (U, A, V, f) \), for \( b \in A, B \subseteq A \) and \( \alpha \in [0, 1] \), \( T^2_b \) and \( T^2_B \) are tolerance relations on \( U \).

**Proof.** For any \( x \in U \), we know \( |b(x) \cap b(x)| = 1 \geq \alpha \). It follows that \((x, x) \in T^2_b \). Hence, \( T^2_b \) is a reflexive relation. Assume \( (x, y) \in T^2_b \), then by **Definition 3.4** we have \( \frac{|b(x) \cap b(y)|}{|b(x) \cup b(y)|} \geq \alpha \). It means \( \frac{|b(y) \cap b(y)|}{|b(y) \cup b(y)|} \geq \alpha \). Consequently, \((y, x) \in T^2_b \). Hence, \( T^2_b \) is a symmetric relation. Then, we know \( T^2_b \) is a tolerance relation on \( U \).

Similarly, we can prove that \( T^2_B \) is a tolerance relation too. □

Let \( (U, A, V, f) \) be a set-valued information system, for \( x \in U, B \subseteq A \), then the indiscernible class of object \( x \) with respect to \( B \) under \( T^2_B \) is defined as

\[ S_{T^2_B}(u) = \{ v \in U \mid (u, v) \in T^2_B \} \]  

(17)

There are two special cases of \( S_{T^2_B} \), i.e. the discrete case and the indiscrete case. The discrete case is defined as

\[ \hat{S}_{T^2_B}(u) = \{ u \}, \forall u \in U \]  

(18)

The indiscrete case is defined as

\[ \tilde{S}_{T^2_B}(u) = U, \forall u \in U \]  

(19)

**Example 3.1.** Table 1 is a set-valued information system \( (U, A, V, f) \), where \( U = \{x_1, x_2, x_3, x_4\} \) and \( A = \{a_1, a_2, a_3\} \). By definition, we have

\[ Q^{0.3}_{a_1} = \{ (x_1, x_1), (x_1, x_2), (x_1, x_3), (x_2, x_1), (x_2, x_2), (x_3, x_1), (x_3, x_2), (x_4, x_1), (x_4, x_2) \} \]

\[ Q^{0.6}_{a_1} = \{ (x_1, x_1), (x_1, x_4), (x_1, x_4), (x_2, x_1), (x_2, x_4), (x_3, x_1), (x_3, x_4), (x_4, x_1), (x_4, x_4) \} \]

\[ Q^{0.4}_{a_1} = \{ (x_1, x_1), (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_1), (x_2, x_2), (x_2, x_3), (x_2, x_4), (x_3, x_1), (x_3, x_3), (x_3, x_4), (x_4, x_1), (x_4, x_4) \} \]

\[ Q^0_{A} = Q^{0.3}_{a_1} \cap Q^{0.6}_{a_1} \cap Q^{0.4}_{a_1} = \{ (x_1, x_1), (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_1), (x_2, x_2), (x_3, x_3), (x_3, x_4), (x_4, x_4) \} \]

Let \( R = Q^0_{A} \), then \( S_R(x_1) = \{ x_1 \}, S_R(x_2) = \{ x_1, x_2 \}, S_R(x_3) = \{ x_3, x_4 \}, S_R(x_4) = \{ x_4 \} \)

**Definition 3.5.** Let \( U \) be a universe, \( P, Q \) be relations on \( U \). We say \( Q \) is coarser than \( P \) (or \( P \) is finer than \( Q \)), denoted by \( P \preceq Q \), if and only if \( \forall u \in U, S_Q(u) \subseteq S_Q(u) \). If \( P \preceq Q \) and \( P \neq Q \) we say \( Q \) is strictly coarser than \( P \) (or \( P \) is strictly finer than \( Q \)), denoted by \( P \prec Q \).

**Proposition 3.3.** In a set-valued information system \( (U, A, V, f) \), for \( b \in A, B \subseteq A \) and \( \alpha \in [0, 1] \), we have \( T^2_b \preceq Q^0_b \) and \( T^2_B \preceq Q^0_B \).

**Proof.** \( \forall y \in S_{T^2_B}(x) \), we have \( \frac{|b(x) \cap b(y)|}{|b(x) \cup b(y)|} \geq \alpha \). Since \( -b(x) \cup b(y) \leq |b(x) \cup b(y)| \), we get \( \frac{|b(x) \cap b(y)|}{|b(x) \cup b(y)|} \geq \alpha \). It follows that \( y \in S_{Q^0}(x) \). Hence, \( T^2_B \preceq Q^0_B \).

Similarly, we can prove that \( T^2_B \preceq Q^0_B \). □

**Theorem 3.1.** Let \( (U, A, V, f) \) be a set-valued information system, \( A_1 \subseteq A_2 \subseteq A \). Let \( A_1 = \{a_1, a_2, \ldots, a_{A_1}\} \) and \( A_2 = \{a_1, a_2, \ldots, a_{A_2}, a_{A_2+1}, \ldots, a_{A_1}\} \). Let \( \alpha = (\alpha_{1}, a_2, \ldots, \alpha_{A_1}) \) and \( \alpha' = (\alpha_{1}, a_2, \ldots, \alpha_{A_1}, \alpha_{A_1+1}, \ldots, \alpha_{A_1}) \). Then, we have \( Q^\alpha_{A_1} \preceq Q^\alpha_{A_1} \) and \( T^\alpha_{A_1} \preceq T^\alpha_{A_1} \).

**Proof.** By **Definition 3.3**, \( \forall u \in U \), we have
For any \( v \in S_{Q_2}^x(u) \), we know \( \forall x_i \in A_1, \bigcap_{i \in A_1} x_i \supseteq x_i \). Since \( x = (x_1, x_2, \ldots, x_{A_1}) \) and \( x' = (x_1, x_2, \ldots, x_{A_1}, x_{A_1+1}, \ldots, x_{A_i}) \), we have \( \forall x_i \in A_1, \bigcap_{i \in A_1} x_i \supseteq x_i \). It means that \( v \in S_{Q_2}^x(u) \). Hence, by Definition 3.5, we know \( Q_2^x \leq Q_1^x \).

Similarly, we can prove \( T_2^x \leq T_1^x \) by Definition 3.4. \( \square \)

**Theorem 3.2.** In a set-valued information system \( \langle U, A, V_f \rangle \), for any attribute \( b \in A \), let \( x_1 \leq x_2 \), then we have \( Q_2^b \leq Q_1^b \) and \( T_2^b \leq T_1^b \).

**Proof.** By Definition 3.3, if \( (x, y) \in Q_2^b \), then \( \frac{|y \cap b|^2}{|x \cap b|^2} \geq x_2 \geq x_1 \). It means that \( (x, y) \in Q_1^b \). It follows that \( \forall x \in U, y \in S_{Q_2^b}(x) \), then \( y \in S_{Q_1^b}(x) \). In other words, \( \forall x \in U, S_{Q_1^b}(x) \subseteq S_{Q_2^b}(x) \). Hence, \( Q_2^b \leq Q_1^b \).

Similarly, we have \( T_2^b \leq T_1^b \) by Definition 3.4. \( \square \)

**Definition 3.6.** For a set-valued information system \( \langle U, A, V_f \rangle \) and \( B \subseteq A \), let \( x = (x_{b_1}, x_{b_2} \ldots x_{b_m}), \forall i, x_{b_i} \in [0, 1] \) and \( \beta = (\beta_{b_1}, \beta_{b_2} \ldots \beta_{b_m}), \forall i, \beta_{b_i} \in [0, 1] \). We say \( x \preceq \beta \), if and only if \( x_{b_i} \leq \beta_{b_i}, \forall b_i \in B \).

**Theorem 3.3.** In a set-valued information system \( \langle U, A, V_f \rangle \), for any attribute set \( B = \{b_1, b_2 \ldots b_m\} \subseteq A \) and \( x = (x_{b_1}, x_{b_2} \ldots x_{b_m}), \forall i, x_{b_i} \in [0, 1] \) and \( \beta = (\beta_{b_1}, \beta_{b_2} \ldots \beta_{b_m}), \forall i, \beta_{b_i} \in [0, 1] \), let \( x \preceq \beta \), then we have \( Q_2^b \leq Q_1^b \) and \( T_2^b \leq T_1^b \).

**Proof.** By Definition 3.3, we have \( Q_2^b = \bigcap_{b_i \in B} Q_2^{b_i} \) and \( Q_1^b = \bigcap_{b_i \in B} Q_1^{b_i} \).

If \( (x, y) \in Q_2^b \), then \( y \in B. (x, y) \in Q_1^b \).

By Definition 3.3, we have \( \forall b_i \in B, \frac{|y \cap b_i|^2}{|x \cap b_i|^2} \geq \beta_{b_i} \geq x_{b_i} \), \( \forall b_i \in B \). It follows that \( \forall y \in B, (x, y) \in Q_1^{b_i} \).

Since \( Q_2^b = \bigcap_{b_i \in B} Q_2^{b_i} \), we have \( (x, y) \in Q_2^b \).

Then, if \( y \in S_{Q_2^b}(x) \), we have \( y \in S_{Q_1^b}(x) \) for all \( x \).

Consequently, \( \forall x \in U, S_{Q_2^b}(x) \subseteq S_{Q_1^b}(x) \). Hence, we have \( Q_2^b \leq Q_1^b \).

\( T_2^b \leq T_1^b \) can be proved similarly by Definition 3.4. \( \square \)

In the following sections, \( R_b \) is used to denote the relation \( Q_2^b \) and \( T_2^b \) for the sake of convenience.

### 4. Entropy measures for set-valued information systems

Yao defined rough entropy in [53] and called it a measure of granularity of a partition. Rough entropy is also called co-entropy by some scholars [3]. Information entropy and rough entropy in single-valued information systems were studied in [25,26]. In this section, definitions of knowledge rough entropy and knowledge information entropy are proposed for set-valued information systems. Their important properties are investigated.

#### 4.1. Knowledge information entropy in set-valued information systems

**Definition 4.1.** Let \( IS = \langle U, A, V_f \rangle \) be a set-valued information system. The information entropy of knowledge A is defined as:

\[
E(A) = \sum_{i=1}^{m} \frac{1}{|U|} \left( 1 - \frac{|S_{R_b}(u_i)|}{|U|} \right)
\]

where \( \frac{|S_{R_b}(u_i)|}{|U|} \) represents the probability of \( S_{R_b}(u_i) \).

By Definitions 3.3 and 3.2, we know both \( Q_1^b \) and \( T_1^b \) are reflexive. It means \( S_{R_b}(u_i) \neq \emptyset \) and \( 1 \leq |S_{R_b}(u_i)| \leq |U| \).

The information entropy of knowledge A achieves the maximum value \( 1 - \frac{1}{|U|} \) when \( \forall i, |S_{R_b}(u_i)| = 1 \), and it achieves the minimum value 1 when \( \forall i, |S_{R_b}(u_i)| = |U| \). Hence, we have \( 0 \leq E(A) \leq 1 - \frac{1}{|U|} \).

**Example 4.1.** (continued from Example 3.1.) Let \( R_b = Q_1^b \), then we have

\[
E(A) = \sum_{i=1}^{m} \frac{1}{|U|} \left( 1 - \frac{|S_{R_b}(u_i)|}{|U|} \right) = \frac{1}{4} \times \left( 1 - \frac{1}{4} \right) + \frac{1}{4} \times \left( 1 - \frac{2}{4} \right) + \frac{1}{4} \times \left( 1 - \frac{1}{4} \right) + \frac{1}{4} \times \left( 1 - \frac{2}{4} \right) = \frac{5}{8}
\]
Theorem 4.1. Let IS = (U,A,V,f) be a set-valued information system and P, Q ⊆ A. If \( R_P < R_Q \) then \( E_i(Q) < E_i(P) \).

Proof. Since \( R_P < R_Q \) by Definition 3.5, we have \( \forall u_i \in U, S_{R_P}(u_i) \subseteq S_{R_Q}(u_i) \) and \( |S_{R_P}(u_i)| \leq |S_{R_Q}(u_i)| \).

Since Q is strictly coarser than P, there must exist \( u_j \in U \) satisfying \( S_{R_Q}(u_j) \subseteq S_{R_P}(u_j) \) and \( |S_{R_Q}(u_j)| = |S_{R_P}(u_j)| \). It follows that

\[
E_i(Q) = -\sum_{i=1}^{\|U\|} \frac{1}{|U|} \log_2 \left( \frac{1}{|S_{R_Q}(u_i)|} \right) < -\sum_{i=1}^{\|U\|} \frac{1}{|U|} \log_2 \left( \frac{1}{|S_{R_P}(u_i)|} \right) = E_i(P).
\]

Hence, the theorem is proved. □

This theorem shows that the knowledge information entropy decreases when the available knowledge becomes coarser, and it increases when available knowledge becomes finer.

In a set-valued information system IS = (U,A,V,f), R is a relation associated with \( x = \{x_b | b \in A\} \). Thus, \( E_i(A) \) is also a function of \( x \) denoted as \( E_i(A)(x) \).

Theorem 4.2. Let IS = (U,A,V,f) be a set-valued information system and \( x \leq \beta \). Then, we have \( E_i(A)(x) \leq E_i(A)(\beta) \).

Proof. Since \( x \leq \beta \), by Theorem 3.3, we have \( Q^x \leq Q^\beta \) and \( T^x \leq T^\beta \). Consequently, we can prove this theorem similar to Theorem 4.1. □

This theorem shows that the knowledge information entropy decreases when the accuracy degree we require gets lower, and it increases when the accuracy degree gets higher.

From Theorems 4.1 and 4.2, we can conclude that the knowledge information entropy introduced in Definition 4.1 can be used to evaluate the certainty degree of knowledge in set-valued information systems. In other words, the more certain the available knowledge is, the bigger the knowledge information entropy value becomes.

4.2. Knowledge rough entropy in set-valued information systems

Definition 4.2. Let IS = (U,A,V,f) be a set-valued information system. The rough entropy of knowledge A is defined as:

\[
E_i(A) = -\sum_{i=1}^{\|U\|} \frac{1}{|U|} \log_2 \left( \frac{1}{|S_{R_P}(u_i)|} \right)
\]

(21)

where \( \frac{1}{|S_{R_P}(u_i)|} \) represents the probability of an element within the class \( S_{R_P}(u_i) \).

The knowledge rough entropy \( E_i(A) \) achieves the maximum value \( \log_2 |U| \) when \( |S_{R_P}(u_i)| = |U|, \forall u_i \in U \).

The knowledge rough entropy \( E_i(A) \) achieves the minimum value 0 when \( |S_{R_P}(u_i)| = 1, \forall u_i \in U \). In other words, when \( S_{R_P}(u_i) = \{u_i\}, \forall u_i \in U \), the rough entropy \( E_i(A) \) achieves the minimum value 0.

Hence, for a set-valued information system IS = (U,A,V,f) we know that \( \log_2 |U| \geq E_i(A) \geq 0 \).

Example 4.2 (Continued from Example 3.1). Let \( R_A = Q^{|0.3,0.6,0.4|}_A \), we have:

\[
E_i(A) = -\sum_{i=1}^{\|U\|} \frac{1}{|U|} \log_2 \left( \frac{1}{|S_{R_A}(u_i)|} \right) = -\frac{1}{4} \times \left( \log_2 \frac{1}{1} + \log_2 \frac{1}{2} + \log_2 \frac{1}{2} + \log_2 \frac{1}{1} \right) = -\frac{1}{4} \times (0 - 1 - 1 + 0) = \frac{1}{2}
\]

Theorem 4.3. Let IS = (U,A,V,f) be a set-valued information system, P, Q ⊆ A. If \( R_P < R_Q \) then we have \( E_i(P) < E_i(Q) \).

Proof. Since \( R_P < R_Q \), by Definition 3.5, we have \( \forall u_i \in U, S_{R_P}(u_i) \subseteq S_{R_Q}(u_i) \) and \( |S_{R_P}(u_i)| \leq |S_{R_Q}(u_i)| \).

Since Q is strictly coarser than P, there must exist \( u_j \in U \) satisfying \( S_{R_Q}(u_j) \subseteq S_{R_P}(u_j) \) and \( |S_{R_Q}(u_j)| = |S_{R_P}(u_j)| \). It follows that

\[
E_i(P) = -\sum_{i=1}^{\|U\|} \frac{1}{|U|} \log_2 \left( \frac{1}{|S_{R_P}(u_i)|} \right) < -\sum_{i=1}^{\|U\|} \frac{1}{|U|} \log_2 \left( \frac{1}{|S_{R_Q}(u_i)|} \right) = E_i(Q).
\]

Hence, the theorem is proved. □

This theorem shows that the knowledge rough entropy increases when the available knowledge becomes coarser, and it decreases when available knowledge becomes finer.

Similarly, in a set-valued information system IS = (U,A,V,f), \( E_i(A) \) is a relation associated with \( x = \{x_b | b \in A\} \). Thus, \( E_i(A) \) is also a function of \( x \) denoted as \( E_i(A)(x) \).

Theorem 4.4. Let IS = (U,A,V,f) be a set-valued information system and \( x \leq \beta \). Then, we have \( E_i(A)(\beta) \leq E_i(A)(x) \).
Proof. Since \( \alpha \leq \beta \), by Theorem 3.3, we have \( Q_\alpha^c \leq Q_\beta^c \) and \( T_\alpha^c \leq T_\beta^c \). Consequently, we can prove this theorem similar to Theorem 4.3. \( \square \)

This theorem shows that the knowledge rough entropy decreases when the accuracy degree we required gets higher, and it increases when the accuracy degree we required gets lower.

From Theorems 4.3 and 4.4, it is safe to conclude that knowledge rough entropy defined in Definition 4.2 can be used to evaluate the uncertainty degree of knowledge in set-valued information systems. In other words, the more uncertain the available knowledge is, the bigger the rough entropy value becomes.

5. Granularity measures for set-valued information systems

The use of entropy of a partition as a measure of granularity was first considered by Wierman in [46]. Knowledge granulation and knowledge granularity measure in single-valued information systems were studied in [25,26]. In this section, definitions of knowledge granulation and granularity measure are proposed for set-valued information systems. Their important properties are also investigated.

5.1. Knowledge granulation in set-valued information systems

**Definition 5.1.** For a set-valued information system \( IS = (U, A, V, f) \). The granulation of knowledge \( A \) is defined as:

\[
GK(A) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |S_{R_A}(u_i)| = \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \left| S_{R_A}(u_i) \right| \right)
\]

(22)

The granulation of knowledge \( A \) achieves the minimum value \( 1/|U| \) when \( |S_{R_A}(u_i)| = 1 \). In other words, \( GK(A) \) achieves the minimum value when \( S_{R_A}(u_i) = \{u_i\}, \forall u_i \in U \).

The granulation of knowledge \( A \) achieves the maximum value \( |U| \) when \( |S_{R_A}(u_i)| = |U| \). In other words, \( GK(A) \) achieves the maximum value when \( S_{R_A}(u_i) = U, \forall u_i \in U \).

Hence, for a set-valued information system, we have \( 1/|U| \geq GK(A) \geq 1 \).

**Example 5.1** (Continued from Example 3.1). Let \( R_A = Q_4^{(0.3,0.6,0.4)} \), then we have

\[
GK(A) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |S_{R_A}(u_i)| = \frac{1}{4^2} \times (1 + 2 + 2 + 1) = \frac{3}{8}
\]

**Theorem 5.1.** Let \( IS = (U, A, V, f) \) be a set-valued information system, and \( P, Q \subseteq A \). If \( R_P \prec R_Q \), then we have \( GK(P) < GK(Q) \).

Proof. Since \( R_P \prec R_Q \) by Definition 3.5, we have \( \forall u_i \in U, S_{R_P}(u_i) \subseteq S_{R_Q}(u_i) \) and \( |S_{R_Q}(u_i)| \leq |S_{R_P}(u_i)| \).

Since \( Q \) is strictly coarser than \( P \), there must exist \( u_i \in U \) satisfying \( S_{R_Q}(u_i) \subset S_{R_P}(u_i) \) and \( |S_{R_P}(u_i)| < |S_{R_Q}(u_i)| \). It follows that

\[
GK(P) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |S_{R_P}(u_i)| < \frac{1}{|U|^2} \sum_{i=1}^{|U|} |S_{R_Q}(u_i)| = GK(Q)
\]

Hence, the theorem is proved. \( \square \)

This theorem shows that the knowledge granulation measure increases when the available knowledge becomes coarser, and it decreases when the available knowledge becomes finer.

Similarly, in a set-valued information system \( IS = (U, A, V, f), E(A) \) is a relation associated with \( \alpha = \{a_b|b \in A\} \). Thus, \( GK(A) \) is also a function of \( \alpha \) denoted as \( GK(A)(\alpha) \).

**Theorem 5.2.** Let \( IS = (U, A, V, f) \) be a set-valued information system. If \( \alpha \leq \beta \), then we have \( GK(A)(\beta) \leq GK(A)(\alpha) \).

Proof. Since \( \alpha \leq \beta \), by Theorem 3.3, we have \( Q_\beta^c \leq Q_\alpha^c \) and \( T_\beta^c \leq T_\alpha^c \). Consequently, we can prove this theorem similar to Theorem 5.1. \( \square \)

This theorem shows that the knowledge granulation measure decreases when the accuracy degree we required gets higher, whereas it increases when the accuracy degree we required gets lower.

From Theorems 5.1 and 5.2, it is concluded that knowledge granulation introduced in Definition 5.1 can be used to evaluate the uncertainty degree of knowledge in set-valued information systems. In other words, the more uncertain the available knowledge is, the bigger the knowledge granulation value becomes.
5.2. Granularity measure in set-valued information systems

Definition 5.2. For a set-valued information system \( IS = \langle U, A, V, f \rangle \), the granularity measure of knowledge \( A \) is defined as:

\[
G(A) = -\frac{1}{|U|} \sum_{i=1}^{n} \log_2 \left( \frac{|S_{R_A}(u_i)|}{|U|} \right)
\]  

(23)

The granularity of knowledge \( A \) achieves the minimum value \( 1/|U| \) when \( |S_{R_A}(u_i)| = 1 \). In other words, \( G(K) \) achieves the minimum value when \( S_{R_A}(u_i) = \{u_i\} \), \( \forall u_i \in U \).

The granularity of knowledge \( A \) achieves the maximum value \( \log_2|U| \) when \( |S_{R_A}(u_i)| = |U| \). In other words, \( G(A) \) achieves the maximum value when \( S_{R_A}(u_i) = U \), \( \forall u_i \in U \).

Hence, for a set-valued information system, we have \( 0 \leq G(A) \leq \log_2|U| \).

Example 5.2 (Continued from Example 3.1). Let \( R_A = Q_{A}^{[0.3,0.6,0.4]} \), then we have

\[
G(A) = -\frac{1}{|U|} \sum_{i=1}^{n} \log_2 \left( \frac{|S_{R_A}(u_i)|}{|U|} \right) = -\frac{1}{4} \times \left( \log_2 \frac{1}{4} + \log_2 \frac{2}{4} + \log_2 \frac{2}{4} + \log_2 \frac{1}{4} \right) = -\frac{1}{4} \times (-2 - 1 - 1 - 2) = \frac{3}{2}
\]

Theorem 5.3. Let \( IS = \langle U, A, V, f \rangle \) be a set-valued information system, and \( P, Q \subseteq A \). If \( R_P \prec R_Q \) then we have \( G(Q) < G(P) \).

Proof. Since \( R_P \prec R_Q \), by Definition 3.5, \( \forall u_i \in U, S_{R_P}(u_i) \subseteq S_{R_Q}(u_i) \) and \( |S_{R_P}(u_i)| \leq |S_{R_Q}(u_i)| \).

Since \( Q \) is strictly coarser than \( P \), there must exist \( u_i \in U \) and \( S_{R_P}(u_i) \subset S_{R_Q}(u_i) \) and \( |S_{R_P}(u_i)| < |S_{R_Q}(u_i)| \).

Consequently, we have

\[
G(Q) = -\frac{1}{|U|} \sum_{i=1}^{n} \log_2 \left( \frac{|S_{R_Q}(u_i)|}{|U|} \right) < -\frac{1}{|U|} \sum_{i=1}^{n} \log_2 \left( \frac{|S_{R_P}(u_i)|}{|U|} \right) = G(P)
\]

This theorem shows that the knowledge granularity measure decreases when the available knowledge becomes coarser, and it increases when the available knowledge becomes finer.

Similarly, in a set-valued information system \( IS = \langle U, A, V, f \rangle \), \( E_i(A) \) is a relation related with \( \alpha = \{x, b \in A \} \). Thus, \( G(A) \) is also a function of \( \alpha \) denoted as \( G(A)(\alpha) \).

Theorem 5.4. Let \( IS = \langle U, A, V, f \rangle \) be a set-valued information system. If \( \alpha \preceq \beta \), then we have \( G(A)(\alpha) \leq G(A)(\beta) \).

Proof. Since \( \preceq \beta \), then we have \( Q_0^b \preceq Q_0^\alpha \) and \( T_0^b \preceq T_0^\alpha \). Then, we can prove this theorem similar to Theorem 5.3. \( \square \)

This theorem shows that the knowledge granularity measure increases when the accuracy degree we required gets higher, whereas it increases when the accuracy degree required gets lower.

From Theorems 5.3 and 5.4, we can conclude that knowledge granularity measure introduced in Definition 5.2 can be used to evaluate the certainty degree of knowledge in set-valued information systems. In other words, the more certain the available knowledge is, the bigger the granularity measure value becomes.

6. Relationship between entropy measures and granulation measures

Theorem 6.1. For an arbitrary set-valued information system \( IS = \langle U, A, V, f \rangle \), we have

\[
G(A) + E_i(A) = \log_2 |U|
\]

Proof. By Definitions 5.2 and 4.2, we have

\[
G(A) = -\frac{1}{|U|} \sum_{i=1}^{n} \log_2 \left( \frac{|S_{R_A}(u_i)|}{|U|} \right) = \frac{1}{|U|} \sum_{i=1}^{n} \left( \log_2 |S_{R_A}(u_i)| - \log_2 |U| \right) = \left( -\frac{1}{|U|} \sum_{i=1}^{n} \log_2 \left( \frac{1}{|S_{R_A}(u_i)|} \right) \right) + \log_2 |U|
\]

Thus, \( G(A) + E_i(A) = \log_2 |U| \). \( \square \)

Example 6.1. For the set-valued information system shown in Table 1, let \( R_A = Q_{A}^{[0.3,0.6,0.4]} \). By Examples 5.2 and 4.2, we have

\[
G(A) + E_i(A) = \frac{3}{2} + \frac{1}{2} = 2 = \log_2 4 = \log_2 |U|
\]
### Theorem 6.2

For an arbitrary set-valued information system \( IS = (U, A, V_f) \), we have

\[
E_i(A) + GK(A) = 1
\]

**Proof.** By Definitions 4.1 and 5.1, we have

\[
E_i(A) = \sum_{i=1}^{\|U\|} \frac{1}{\|U\|} \left( 1 - \frac{|S_{E_i}(u_i)|}{|U|} \right) = \sum_{i=1}^{\|U\|} \frac{1}{|U|} \sum_{i=1}^{\|U\|} \frac{|S_{E_i}(u_i)|}{|U|^2} = 1 - GK(A) \quad \square
\]

### Example 6.2

For the set-valued information system shown in Table 1, let \( R_A = Q_A^{[0.3,0.6,0.4]} \), by Examples 4.1 and 5.1, we have:

\[
E_i(A) + GK(A) = \frac{5}{8} + \frac{3}{8} = 1
\]

It should be pointed out that a similar type of connection to Theorem 6.1 was considered by Wierman [46] and Yao [53] implicitly under the framework of partition. Similar types of connection to Theorems 6.1 and 6.2 in single-valued information systems were studied in [25,26].

### 7. Conclusion

In this paper, we define two new relations for set-valued information systems. The concepts of knowledge information entropy, knowledge rough entropy, knowledge granulation and knowledge granularity measure are defined and studied in set-valued information systems based on the proposed relations. It is shown that knowledge information entropy and knowledge granularity measure can be used to evaluate the certainty degree of knowledge in set-valued information systems, and knowledge rough entropy and knowledge granulation can be used to evaluate the uncertainty degree of knowledge in set-valued information systems. Fuzzy rough set models for set-valued information systems will be considered and further explored in our future study.

### Acknowledgements

The authors are grateful to the reviewers and Associate Editor for their valuable suggestions in improving the paper. This work is supported in part by the National Natural Science Foundation of China (Nos. 61070074 and 60703038).

### References
