Science in China Series F: Information Sciences

© 2008 SCIENCE IN CHINA PRESS

Springer

www.scichina.com
info.scichina.com
www.springerlink.com

Information granules and entropy theory in information systems

LIANG JiYe† & QIAN YuHua

Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education; School of Computer & Information Technology, Shanxi University, Taiyuan 030006, China

Information granulation and entropy theory are two main approaches to research uncertainty of an information system, which have been widely applied in many practical issues. In this paper, the characterizations and representations of information granules under various binary relations are investigated in information systems, an axiom definition of information granulation is presented, and some existing definitions of information granulation become its special forms. Entropy theory in information systems is further developed and the granulation monotonicity of each of them is proved. Moreover, the complement relationship between information granulation and entropy is established. This investigation unifies the results of measures for uncertainties in complete information systems and incomplete information systems.

information systems, information granule, information granulation, entropy, rough set

1 Introduction

Granular computing (GrC) was presented by Zadeh in 1996[1]. He identified three basic concepts that underline the process of human cognition, namely, granulation, organization, and causation. "Granulation involves decomposition of whole into parts, organization involves integration of parts into whole, and causation involves association of causes and effects". GrC is an umbrella term to cover any theories, methodologies, techniques, and tools that make use of granules in problem solving[2−4]. As an effective tool for complex problem solving, it has potential applications in rough set theory, concept lattice, knowledge engineering, data mining, artificial intelligence, machine learning, etc., and has become an important research issue in information sciences field[5−12].

Granular computing has three important models: 1) computing with words[1, 2], 2) rough set

Received November 6, 2006; accepted December 26, 2007
doi: 10.1007/s11432-008-0113-2
†Corresponding author (email: ljy@sxu.edu.cn)
Supported by the National Natural Science Foundation of China (Grant No. 60773133), the National Key Basic Research and Pevelopment Program of China (973) (Grant No. 2007CB311002), and the National 863 Project (Grant No. 2007AA01Z165)

theory[13], and 3) quotient space theory[14−16].

Information granulation is mainly used to study uncertainty of information or knowledge in information systems. Wierman[17] introduced the concept of granulation measure to measure uncertainty of information. This concept has the same form as Shannon's entropy under the axiom definition presented. Liang et al.[18−20] investigated information granulation in complete/incomplete information systems, which have been effectively applied in measuring for attribute significance, feature selection, decision-rule extracting, etc. Qian and Liang[21] introduced combination granulation with intuitionistic knowledge-content nature to measure the size of information granulation in information systems.

The concept of entropy that comes from classical energetics can be used to measure out-of-order degree of a system. The entropy of a system as defined by Shannon[22], called information entropy, gives a measure of uncertainty about its actual structure. It has been a useful mechanism for characterizing uncertainty in various modes and applications in many diverse fields. Düntsch[23] used Shannon's entropy to measure for decision rules in rough set theory. In refs. [24, 25], Beaubouef and Petry introduced a variant of Shannon's entropy to investigate uncertainty measure in rough set theory and rough relationship database. In 1968, Zadeh defined the entropy of a fuzzy set as a weighted Shannon's entropy. It is the first attempt for measuring fuzzy information. In 1972, De Luca and Termini[26] formulated the axioms of fuzzy entropy, gave a distance measure and similarity measure of fuzziness of fuzzy sets, and established the relationship between these two concepts. In ref. [27], the author gave a new fuzzy entropy and applied it for measuring for the fuzziness of a fuzzy-rough set based partition. A new information entropy was proposed by Liang in ref. [28] and its conditional entropy and mutual information were defined. Unlike the logarithmic behavior of Shannon's entropy, the gain function of Liang's entropy possesses the complement nature. This entropy can be used to measure the fuzziness of a rough set and a rough classification. Qian and Liang[21] proposed combination entropy with intuitionistic knowledge-content characteristic in incomplete information systems, which can be used to measure the uncertainty of an incomplete information system.

In this paper, we unify some measures for uncertainties in complete information systems and incomplete information systems. By introducing a partial relation $\preceq'$, an axiom definition of information granulation is given in information systems and some existing measures of information granulation are all its special forms. Entropy theory is further developed in incomplete information systems, granulation monotonicity of each of the entropies is proved, and the complement relationship between information granulation and entropy is established. These results provide effectual tools for studying uncertainty in information systems.

2 Information granules in information systems

2.1 Complete information systems

An information system is pair $S=(U,A)$, where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, and $a: U \to V_a$ is a mapping for $a \in A$, where V_a is called the value set of a.

For an information system $S=(U,A)$, if $a \in A$ and every element in V_a is a definite value, then S is called a complete information system[19,20,29].

Let $P \subseteq A$, define an equivalence relation

$$IND(P) = \{(u,v) \in U \times U \mid \forall a \in P, a(u) = a(v)\}.$$

Obviously, $IND(P) = \bigcap_{a \in P} IND(\{a\})$.

$U / IND(P)$ constitutes a partition of U. $U / IND(P)$ is called an information on U, and every equivalence class is called an information granule. Information granulation denotes average measure of information granules (equivalence classes) under attributes set P.

In particular, if $U / IND(P) = \omega = \{X \mid X = \{u\}, u \in U\}$, it is called identity relation; if $U / IND(P) = \delta = \{X \mid X = \{U\}\}$, it is called universal relation.

2.2 Incomplete information systems

It may happen that some of the attribute values for an object are missing. For example, in medical information systems, there may exist a group of patients for which it is impossible to perform all the required tests. These missing values can be represented by the set of all possible values for the attribute or equivalence by the domain of the attribute. To indicate such a situation, a distinguished value, a so-called null value, is usually assigned to those attributes.

In the information system $S = (U, A)$, if there at least exists an attribute $a \in A$ such that V_a contains a null value, then S is called an incomplete information system. Further on, we will denote the null value by *[19,20,29].

Let $P \subseteq A$, define a tolerance relation[29]:

$$SIM(P) = \{(u,v) \in U \times U \mid \forall a \in P, a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *\}.$$

Obviously, $SIM(P) = \bigcap_{a \in P} SIM(\{a\})$.

Let $S_P(u)$ denote the object set $\{v \in U \mid (u,v) \in SIM(P)\}$ and $S_P(u)$ is the tolerance class of u relative to P, called an information granule. Let $\tilde{P}$ represent the classification, i.e., the family sets $\{S_P(u) \mid u \in U\}$. $\tilde{P}$ constitutes a covering of U, i.e., $S_P(u) \neq \varnothing$ for every $u \in U$ and $\bigcup_{u \in U} S_P(u) = U$. Here, $\tilde{P}$ is called an information on U and every tolerance class is called an information granule. Information granulation denotes average measure of information granules (tolerance classes) under attributes set P.

In particular, if $\tilde{P} = \omega = \{S_P(u) \mid S_P(u) = \{u\}, u \in U\}$, it is called an identity relation; if $\tilde{P} = \delta = \{S_P(u) \mid S_P(u) = \{U\}, u \in U\}$, it is called a universal relation.

Example 1. Consider the descriptions of several cars in Table 1[20,29].

Table 1 An incomplete information system about cars[20,29]

Car	Price	Mileage	Size	Max-speed
u_1	high	low	full	low
u_2	low	*	full	low
u_3	*	*	compact	low
u_4	high	*	full	high
u_5	*	*	full	high
u_6	low	high	full	*

This is an incomplete information system, where $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$, $A = \{a_1, a_2, a_3, a_4\}$.

For convenience, we denote Price, Mileage, Size, and Max-speed by a_1, a_2, a_3, and a_4, respectively. By computing, it follows that

$$\tilde{A}=\{S_A(u_1),S_A(u_2),S_A(u_3),S_A(u_4),S_A(u_5),S_A(u_6)\},$$

where $S_A(u_1)=\{u_1\}$, $S_A(u_2)=\{u_2,u_6\}$, $S_A(u_3)=\{u_3\}$, $S_A(u_4)=\{u_4,u_5\}$, $S_A(u_5)=\{u_4,u_5,u_6\}$, and $S_A(u_6)=\{u_2,u_5,u_6\}$.

Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$, then $\tilde{A}=\{S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|})\}$ [20,21]. Let the set $X_i=\{u_{i1},u_{i2},\cdots,u_{is_i}\}$, where $|X_i|=s_i$ and $\sum_{i=1}^{m}s_i=|U|$, then

$$X_i=S_A(u_{i1})=S_A(u_{i2})=\cdots=S_A(u_{is_i}),\quad |X_i|^2=\sum_{k=1}^{s_i}|S_A(u_{ik})|.$$

Hence, when we do not need to distinguish complete information systems and incomplete information systems, an information in an information system can be represented as the vector $K(A)=(S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|}))$ [20,21]. In particular, the identity relation can be denoted by the vector $\tilde{P}=\omega=(\{u_1\},\{u_2\},\cdots,\{u_{|U|}\})$ and the universal relation can be denoted by the vector $\tilde{P}=\delta=(\{U\},\{U\},\cdots,\{U\})$.

Let $S=(U,A)$ be a complete information system, $P,Q\subseteq A$, $U/IND(P)=\{X_1,X_2,\cdots,X_m\}$, $U/IND(Q)=\{Y_1,Y_2,\cdots,Y_n\}$, $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$, and $K(Q)=(S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|}))$, where $X_i=\{u_{i1},u_{i2},\cdots,u_{is_i}\}$ and $Y_j=\{u_{j1},u_{j2},\cdots,u_{jt_j}\}$. Then, the relationship between $K(P)$ and $U/IND(P)$ is as follows:

$$X_i=S_P(u_{i1})=S_P(u_{i2})=\cdots=S_P(u_{is_i}),\quad |X_i|^2=\sum_{k=1}^{s_i}|S_P(u_{ik})|.$$

Similarly, the relationship between $K(Q)$ and $U/IND(Q)$ is as follows:

$$Y_j=S_Q(u_{j1})=S_Q(u_{j2})=\cdots=S_Q(u_{jt_j}),\quad |Y_j|^2=\sum_{k=1}^{t_j}|S_Q(u_{jk})|.$$

These kinds of representation forms will be largely used in the proofs of many theorems in this paper.

3 Axiom approach of information granulation

3.1 Partial relation "$\preceq'$"

Let $S=(U,A)$ be an information system, $P,Q\subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$, and $K(Q)=(S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|}))$. We define a binary relation $\preceq'$:

$$K(P)\preceq' K(Q)\Leftrightarrow \text{if any } i\in\{1,2,\cdots,|U|\},$$

one has $S_P(u_i)\subseteq S_Q(u_i)$, where $S_P(u_i)\in K(P)$ and $S_Q(u_i)\in K(Q)$, just $P\preceq' Q$.

Further on, $K(P)=K(Q) \Leftrightarrow$ if any $i \in \{1,2,\cdots,|U|\}$, one has $S_P(u_i)=S_Q(u_i)$, just $P \approx Q$; $K(P) \prec' K(Q) \Leftrightarrow K(P) \preceq' K(Q)$ and $K(P) \neq K(Q)$, just $P \prec' Q$.

Let $S=(U,A)$ be an information system, denoted by $K=\{K(P) \mid P \subseteq A\}$.

Theorem 1. $(K,\preceq')$ is a partial set.

Proof. Let $P,Q,R \subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$, $K(Q)=(S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|}))$ and $K(R)=(S_R(u_1),S_R(u_2),\cdots,S_R(u_{|U|}))$.

1) It is obvious that $|S_P(u)|=|S_P(u)|$, $u \in U$, hence $P \preceq' P$.

2) Suppose $P \preceq' Q$ and $Q \preceq' P$. From the above definitions, we can obtain that

$P \preceq' Q \Leftrightarrow S_P(u_i) \subseteq S_Q(u_i)$, $i \in \{1,2,\cdots,|U|\}$, where $S_P(u_i) \in K(P)$ and $S_Q(u_i) \in K(Q)$;

$Q \preceq' P \Leftrightarrow S_Q(u_i) \subseteq S_P(u_i)$, $i \in \{1,2,\cdots,|U|\}$, where $S_Q(u_i) \in K(Q)$ and $S_P(u_i) \in K(P)$.

Therefore, we have $S_P(u_i) \subseteq S_Q(u_i) \subseteq S_P(u_i)$, i.e., $S_P(u_i)=S_Q(u_i)$.

Hence, we have $S_P(u_i)=S_Q(u_i)$, $\forall u_i \in U$, i.e., $P \approx Q$.

3) Suppose $P \preceq' Q$ and $Q \preceq' R$. From the above definitions, we can obtain that

$P \preceq' Q \Leftrightarrow S_P(u_i) \subseteq S_Q(u_i)$, $i \in \{1,2,\cdots,|U|\}$, where $S_P(u_i) \in K(P)$ and $S_Q(u_i) \in K(Q)$;

$Q \preceq' R \Leftrightarrow S_Q(u_i) \subseteq S_R(u_i)$, $i \in \{1,2,\cdots,|U|\}$, where $S_Q(u_i) \in K(Q)$ and $S_R(u_i) \in K(R)$.

Thus, we have $S_P(u_i) \subseteq S_Q(u_i) \subseteq S_R(u_i)$, $i \in \{1,2,\cdots,|U|\}$, i.e., $S_P(u_i) \subseteq S_R(u_i)$. Hence, $P \preceq' R$.

Therefore, $(K,\preceq')$ is a partial set.

3.2 Axiom definition of information granulation and its properties

Definition 1. Let $S=(U,A)$ be an information system, $P \subseteq A$ and G be a mapping from the power set of A to the set of real numbers. We say that G is an information granulation in an information system if G satisfies the following conditions:

1) $G(P) \geqslant 0$; (Non-negativity)

2) $\forall P,Q \subseteq A$, if $P \approx Q$, then $G(P)=G(Q)$; (Invariability)

3) $\forall P,Q \subseteq A$, if $P \prec' Q$, then $G(P)<G(Q)$. (Monotonicity)

Theorem 2 (Extremum). Let $S=(U,A)$ be an information system and $P \subseteq A$, then if $K(P)=\omega$, $G(P)$ achieves its minimum value; if $K(P)=\delta$, $G(P)$ achieves its maximum value.

Proof. Let $K(\omega)=(\{u_1\},\{u_2\},\cdots,\{u_{|U|}\})$ and $K(\delta)=(\{U\},\{U\},\cdots,\{U\})$. Hence, for any $R \subseteq A$, $K(R)=(S_R(u_1),S_R(u_2),\cdots,S_R(u_{|U|}))$, we have $\{u_i\} \subseteq S_R(u_i)$ $(u_i \in U)$, i.e., $\omega \preceq' R$. Similarly, we have $S_R(u_i) \subseteq U$ $(u_i \in U)$, i.e., $R \preceq' \delta$.

From (2) and (3) in Definition 1, we know that $G(\omega) \leqslant G(R) \leqslant G(\delta)$. QED

In the following, we prove that some of existing definitions are all various special forms of information granulation.

Definition 2[18]. Let $S=(U,A)$ be a complete information system, $U/IND(A)=\{X_1, X_2, \cdots, X_m\}$. Information granulation of A is defined as

$$GK(A)=\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2, \tag{1}$$

where $\sum_{i=1}^{m}|X_i|^2$ is the number of elements in the equivalence relation induced by $\bigcup_{i=1}^{m}(X_i \times X_i)$.

If $U/IND(A)=\omega$, then the information granulation of A achieves its minimum value $|U|/|U|^2=1/|U|$.

If $U/IND(A)=\delta$, then the information granulation of A achieves its maximum value $|U|^2/|U|^2=1$.

Theorem 3. *GK* in Definition 2 is an information granulation under Definition 1.

Proof.

1) Obviously, its non-negativity holds.

2) Let $P,Q \subseteq A$, , then the information $U/IND(P)=\{X_1, X_2, \cdots, X_m\}$ and the information $U/IND(Q)=\{Y_1, Y_2, \cdots, Y_n\}$ in complete information systems can be represented as the information $K(P)=(S_P(u_1), S_P(u_2), \cdots, S_P(u_{|U|}))$ and the information $K(Q)=(S_Q(u_1), S_Q(u_2), \cdots, S_Q(u_{|U|}))$, respectively.

If $P \approx Q$, then we have $S_P(u_i)=S_Q(u_i)$, $i \in \{1,2,\cdots,|U|\}$, i.e., $|S_P(u_i)|=|S_Q(u_i)|$. Therefore, we have that

$$GK(P)=\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=\frac{1}{|U|^2}\sum_{j=1}^{n}|Y_j|^2=GK(Q).$$

3) Let $P,Q \subseteq A$ with $P \prec' Q$, then we have that $S_P(u_i) \subseteq S_Q(u_i)$ and $|S_P(u_i)| \leqslant |S_Q(u_i)|$, $u_i \in U$, and there exists $u_0 \in U$ such that $S_P(u_0) \subset S_Q(u_0)$ and $|S_P(u_0)| < |S_Q(u_0)|$. Therefore,

$$\begin{aligned}GK(P)&=\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|=\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_P(u_i)|+|S_P(u_0)|\right)\\&<\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_Q(u_i)|+|S_Q(u_0)|\right)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=\frac{1}{|U|^2}\sum_{j=1}^{n}|Y_j|^2=GK(Q).\end{aligned}$$

Hence, *GK* in Definition 2 is an information granulation under Definition 1.

Definition 3[20]. Let $S=(U,A)$ be an incomplete information system and $K(A)=(S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|}))$. Information granulation of A is defined as

$$GK(A)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_A(u_i)|. \tag{2}$$

If $K(A)=\omega$, then the information granulation of A achieves its minimum value $|U|/|U|^2=1/|U|$.

If $K(A)=\delta$, then the information granulation of A achieves its maximum value $|U|^2/|U|^2=1$.

Theorem 4. GK in Definition 3 is an information granulation under Definition 1.

Proof.

1) Obviously, its non-negativity holds.

2) Let $P,Q\subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$ and $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|U|}))$.

If $P\approx Q$, we have that $S_P(u_i)=S_Q(u_i)$, $i\in\{1,2,\cdots,|U|\}$, i.e., $|S_P(u_i)|=|S_Q(u_i)|$. Hence, we can obtain that

$$GK(P)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=GK(Q).$$

3) Let $P,Q\subseteq A$ with $P\prec' Q$, we know that $S_P(u_i)\subseteq S_Q(u_i)$ and $|S_P(u_i)|\leqslant|S_Q(u_i)|$, $u_i\in U$, and there exists $u_0\in U$ such that $S_P(u_0)\subset S_Q(u_0)$ and $|S_P(u_0)|<|S_Q(u_0)|$. Thus,

$$\begin{aligned}GK(P)&=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|=\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_P(u_i)|+|S_P(u_0)|\right)\\&<\frac{1}{|U|^2}\left(\sum_{i=1,u_i\neq u_0}^{|U|}|S_Q(u_i)|+|S_Q(u_0)|\right)=\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|=GK(Q).\end{aligned}$$

Therefore, GK in Definition 3 is an information granulation under Definition 1.

Definition 4[21]. Let $S=(U,A)$ be a complete information system, $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. Combination granulation of A is defined as

$$CG(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2}, \tag{3}$$

where $\frac{|X_i|}{|U|}$ represents the probability of equivalence class X_i within the universe U; $\frac{C_{|X_i|}^2}{C_{|U|}^2}$ denotes the probability of pairs of elements on equivalence class X_i within the whole pairs of elements on the universe U.

If $U/IND(A)=\omega$, then the combination granulation of A achieves its minimum value $0/C_{|U|}^2=0$.

If $U/IND(A)=\delta$, then the combination granulation of A achieves its maximum value $C_{|U|}^2/C_{|U|}^2=1$.

Theorem 5. CG in Definition 4 is an information granulation under Definition 1.

Proof.

1) Obviously, its non-negativity holds.

2) Let $P,Q\subseteq A$, then the information $U/IND(P)=\{X_1,X_2,\cdots,X_m\}$ and the information $U/IND(Q)=\{Y_1,Y_2,\cdots,Y_n\}$ in complete information systems can be represented as the information $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$ and the information $K(Q)=(S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|}))$, respectively. Suppose $X_i=\{u_{i1},u_{i2},\cdots,u_{is_i}\}$, $i\in\{1,2,\cdots,m\}$, where $|X_i|=s_i$ and $\sum_{i=1}^{m}s_i=|U|$, then the relationship between $K(P)$ and $U/IND(P)$ is as follows:

$$X_i=S_P(u_{i1})=S_P(u_{i2})=\cdots=S_P(u_{is_i}),$$

$$|X_i|=|S_P(u_{i1})|=|S_P(u_{i2})|=\cdots=|S_P(u_{is_i})|=s_i.$$

Similarly, $|Y_j|=|S_Q(u_{j1})|=|S_Q(u_{j2})|=\cdots=|S_Q(u_{jt_j})|=t_j$.

Therefore, we have that

$$|X_i|\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=s_i\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=\sum_{k=1}^{s_i}\frac{C_{|S_P(u_{ik})|}^2}{C_{|U|}^2},$$

$$|Y_j|\times\frac{C_{|Y_j|}^2}{C_{|U|}^2}=t_j\times\frac{C_{|Y_j|}^2}{C_{|U|}^2}=\sum_{k=1}^{t_j}\frac{C_{|S_Q(u_{jk})|}^2}{C_{|U|}^2}.$$

If $P\approx Q$, one has $S_P(u_i)=S_Q(u_i)$, $i\in\{1,2,\cdots,|U|\}$, i.e., $|S_P(u_i)|=|S_Q(u_i)|$. Hence, we can obtain that

$$CG(P)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|X_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{m}s_i\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=\frac{1}{|U|}\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{C_{|S_P(u_{ik})|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}$$

$$=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}=\sum_{j=1}^{n}\sum_{k=1}^{t_j}\frac{C_{|S_Q(u_{jk})|}^2}{C_{|U|}^2}=\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\frac{C_{|Y_j|}^2}{C_{|U|}^2}=CG(Q).$$

3) Let $P,Q\subseteq A$ with $P\prec' Q$, then we have $S_P(u_i)\subseteq S_Q(u_i)$ and $|S_P(u_i)|\leqslant|S_Q(u_i)|$, $u_i\in U$, and there exists $u_0\in U$ such that $S_P(u_0)\subset S_Q(u_0)$ and $|S_P(u_0)|<|S_Q(u_0)|$. Hence,

$$CG(P)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{X_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}=\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\frac{C_{S_P(u_i)|}^2}{C_{|U|}^2}+\frac{1}{|U|}\frac{C_{|S_P(u_0)|}^2}{C_{|U|}^2}$$

$$<\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}+\frac{1}{|U|}\frac{C_{|S_Q(u_0)|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}=\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\frac{C_{|Y_j|}^2}{C_{|U|}^2}=CG(Q).$$

Thus, CG in Definition 4 is an information granulation under Definition 1.

Definition 5[21]**.** Let $S=(U,A)$ be an incomplete information system and $K(A)=(S_A(u_1),$ $S_A(u_2),\cdots,S_A(u_{|U|}))$. Combination granulation of A is defined as

$$CG(A)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|S_A(u_i)|}}{C^2_{|U|}}, \tag{4}$$

where $\frac{C^2_{|S_A(u_i)|}}{C^2_{|U|}}$ denotes the probability of pairs of elements on the tolerance class $S_A(u_i)$ within the whole pairs of elements on the universe U.

If $K(A)=\omega$, then the combination granulation of A achieves its minimum value $|U|\times 0/|U|=0$.

If $K(A)=\delta$, then the combination granulation of A achieves its maximum value $|U|/|U|=1$.

Theorem 6. CG in Definition 5 is an information granulation under Definition 1.

Proof.

1) Obviously, its non-negativity holds.

2) Let $P,Q\subseteq A$, $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$ and $K(Q)=(S_Q(u_1),S_Q(u_2),\ldots,S_Q(u_{|U|}))$.

If $P\approx Q$, we know that $S_P(u_i)=S_Q(u_i)$, $i\in\{1,2,\cdots,|U|\}$, i.e., $|S_P(u_i)|=|S_Q(u_i)|$. Thus, we have that

$$CG(P)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}}=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|S_Q(u_i)|}}{C^2_{|U|}}=CG(Q).$$

3) Let $P,Q\subseteq A$ with $P\prec' Q$, we have that $S_P(u_i)\subseteq S_Q(u_i)$ and $|S_P(u_i)|\leqslant|S_Q(u_i)|$, $u_i\in U$, and there exists $u_0\in U$ such that $S_P(u_0)\subset S_Q(u_0)$ and $|S_P(u_0)|<|S_Q(u_0)|$. Hence,

$$\begin{aligned}CG(P)&=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}}=\sum_{i=1,u_i\neq u_0}^{|U|}\frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}}+\frac{C^2_{|S_P(u_0)|}}{C^2_{|U|}}\\&<\sum_{i=1,u_i\neq u_0}^{|U|}\frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}}+\frac{C^2_{|S_P(u_0)|}}{C^2_{|U|}}=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|S_P(u_i)|}}{C^2_{|U|}}=CG(Q).\end{aligned}$$

Therefore, CG in Definition 5 is an information granulation under Definition 1.

Definition 6[18,20]. Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. Rough entropy of A is defined as

$$E_r(A)=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}. \tag{5}$$

If $U/IND(A)=\omega$, then the rough entropy of A achieves its minimum value 0.

If $U/IND(A)=\delta$, then the rough entropy of A achieves its maximum value $\log_2|U|$.

Theorem 7. E_r in Definition 6 is an information granulation under Definition 1.

Proof.

1) Obviously, its non-negativity holds.

2) Let $P,Q\subseteq A$, $U/IND(P)=\{X_1,X_2,\cdots,X_m\}$, $U/IND(Q)=\{Y_1,Y_2,\cdots,Y_n\}$ and $K(P)=$

$(S_P(u_1), S_P(u_2), \cdots, S_P(u_{|U|}))$, $K(Q) = (S_Q(u_1), S_Q(u_2), \cdots, S_Q(u_{|U|}))$. If $P \approx Q$, we have that $S_P(u_i) = S_Q(u_i)$, $i \in \{1, 2, \cdots, |U|\}$, i.e., $|S_P(u_i)| = |S_Q(u_i)|$. Therefore,

$$E_r(P) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|} = -\sum_{i=1}^{m}\sum_{k=1}^{s_i} \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_{ik})|} = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_i)|}$$

$$= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_Q(u_i)|} = -\sum_{j=1}^{n} \frac{|Y_j|}{|U|} \log_2 \frac{1}{|Y_j|} = E_r(Q).$$

3) Let $P, Q \subseteq A$ with $P \prec' Q$, we have that $S_P(u_i) \subseteq S_Q(u_i)$ and $1 \leqslant |S_P(u_i)| \leqslant |S_Q(u_i)|$, $u_i \in U$, and there exists $u_0 \in U$ such that $S_P(u_0) \subset S_Q(u_0)$ and $1 \leqslant |S_P(u_0)| < |S_Q(u_0)|$. Hence,

$$\begin{aligned}
E_r(P) &= -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|} \\
&= -\sum_{i=1}^{m} \left(\frac{1}{|U|} \log_2 \frac{1}{|S_P(u_{i1})|} + \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_{i2})|} + \cdots + \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_{is_i})|} \right) \\
&= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_i)|} = \sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 |S_P(u_i)| \\
&= \frac{1}{|U|} \sum_{i=1, u_i \neq u_0}^{|U|} \log_2 |S_P(u_i)| + \frac{1}{|U|} \log_2 |S_P(u_0)| \\
&= \frac{1}{|U|} \log_2 \prod_{i=1, u_i \neq u_0}^{|U|} |S_P(u_i)| + \frac{1}{|U|} \log_2 |S_P(u_0)| \\
&< \frac{1}{|U|} \log_2 \prod_{i=1, u_i \neq u_0}^{|U|} |S_Q(u_i)| + \frac{1}{|U|} \log_2 |S_Q(u_0)| \\
&= \frac{1}{|U|} \log_2 \prod_{i=1}^{|U|} |S_Q(u_i)| = \sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 |S_Q(u_i)| \\
&= -\sum_{j=1}^{n}\sum_{k=1}^{t_j} \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_{jk})|} = -\sum_{j=1}^{n} \frac{|Y_j|}{|U|} \log_2 \frac{1}{|Y_j|} \\
&= E_r(Q),
\end{aligned}$$

that is $E_r(P) < E_r(Q)$.

Thus, E_r in Definition 6 is an information granulation under Definition 1.

Definition 7[19]. Let $S = (U, A)$ be an incomplete information system and $K(A) = (S_A(u_1), S_A(u_2), \cdots, S_A(u_{|U|}))$. Rough entropy of A is defined as

$$E_r(A) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_i)|}. \tag{6}$$

If $K(A) = \omega$, then the rough entropy of A achieves its minimum value 0.

If $K(A) = \delta$, then the rough entropy of A achieves its maximum value $\log_2 |U|$.

Theorem 8. Let $S = (U, A)$ be a complete information system and $U / IND(A) =$

$\{X_1, X_2, \cdots, X_m\}$. Then, the rough entropy of A degenerates into

$$E_r(A) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|}.$$

Proof. Suppose that $K(A) = (S_A(u_1), S_A(u_2), \cdots, S_A(u_{|U|}))$ and $X_i = \{u_{i1}, u_{i2}, \cdots, u_{is_i}\}$, where $|X_i| = s_i$ and $\sum_{i=1}^{m} s_i = |U|$. Then, the relationship between them is as follows

$$X_i = S_A(u_{i1}) = S_A(u_{i2}) = \cdots = S_A(u_{is_i}).$$

Therefore,

$$\begin{aligned}
&-\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|} \\
&= -\sum_{i=1}^{m} \left(\frac{1}{|U|} \log_2 \frac{1}{|S_A(u_{i1})|} + \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_{i2})|} + \cdots + \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_{is_i})|} \right) \\
&= -\left(\frac{1}{|U|} \log_2 \frac{1}{|S_A(u_1)|} + \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_2)|} + \cdots + \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_{|U|})|} \right) \\
&= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_i)|} \\
&= E_r(A).
\end{aligned}$$

This completes the proof.

Theorem 9. E_r in Definition 7 is an information granulation under Definition 1.

Proof. The proof is similar to that of Theorem 7.

4 Entropy theory in information systems

Entropy is always used to measure out-of-order degree of a system. The bigger entropy value is, the higher our-of-order of a system. Shannon introduced the concept of entropy in physics to information theory for measuring uncertainty of the structure of a system.

Definition 8[22]. Let $S = (U, A)$ be a complete information system, $U/IND(A) = \{X_1, X_2, \cdots, X_m\}$ and $p_i = \frac{|X_i|}{|U|}$, we call

$$H(A) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{7}$$

information entropy of the information system S. When some p_i equals 0, then $0 \cdot \log_2 0 = 0$.

Theorem 10. Let $S = (U, A)$ be a complete information system and $P, Q \subseteq A$. If $P \prec' Q$, then $H(Q) < H(P)$.

Proof. Let $U/IND(P) = \{X_1, X_2, \cdots, X_m\}$ and $U/IND(Q) = \{Y_1, Y_2, \cdots, Y_n\}$. Suppose that $X_i = \{u_{i1}, u_{i2}, \cdots, u_{is_i}\}$ and $Y_j = \{u_{j1}, u_{j2}, \cdots, u_{jt_j}\}$, where $|X_i| = s_i$, $|Y_j| = t_j$, and $\sum_{i=1}^{m} s_i = |U|$,

$\sum_{j=1}^{n} t_j = |U|$. They can be uniformly expressed as $K(P) = (S_P(u_1), S_P(u_2), \cdots, S_P(u_{|U|})\}$ and $K(Q) = \{S_Q(u_1), S_Q(u_2), \cdots, S_Q(u_{|U|}))$. Hence, the relationship between $K(P)$ and $U/IND(P)$ can be established as follows:

$$X_i = S_P(u_{i1}) = S_P(u_{i2}) = \cdots = S_P(u_{is_i}),$$
$$|X_i| = |S_P(u_{i1})| = |S_P(u_{i2})| = \cdots = |S_P(u_{is_i})| = s_i.$$

Similarly, we have $|Y_j| = |S_Q(u_{j1})| = |S_Q(u_{j2})| = \cdots = |S_Q(u_{jt_j})| = t_j$.

Therefore,

$$\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|} = \frac{1}{|U|}\log_2\frac{|S_P(u_{i1})|}{|U|} + \frac{1}{|U|}\log_2\frac{|S_P(u_{i2})|}{|U|} + \cdots + \frac{1}{|U|}\log_2\frac{|S_P(u_{is_i})|}{|U|},$$
$$\frac{|Y_j|}{|U|}\log_2\frac{|Y_j|}{|U|} = \frac{1}{|U|}\log_2\frac{|S_Q(u_{j1})|}{|U|} + \frac{1}{|U|}\log_2\frac{|S_Q(u_{j2})|}{|U|} + \cdots + \frac{1}{|U|}\log_2\frac{|S_Q(u_{jt_j})|}{|U|}.$$

Since $P \prec' Q$, thus one has $S_P(u_i) \subseteq S_Q(u_i)$ and $1 \leqslant |S_P(u_i)| \leqslant |S_Q(u_i)|$, $u_i \in U$, and there exists $u_0 \in U$ such that $S_P(u_0) \subset S_Q(u_0)$ and $1 \leqslant |S_P(u_0)| < |S_Q(u_0)|$.

Hence,

$$\begin{aligned}
H(P) &= -\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|} \\
&= -\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\log_2\frac{|S_P(u_{ik})|}{|U|} = -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)|}{|U|} \\
&= -\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)|}{|U|} - \frac{1}{|U|}\log_2\frac{|S_P(u_0)|}{|U|} \\
&> -\sum_{i=1,u_i\neq u_0}^{|U|}\frac{1}{|U|}\log_2\frac{|S_Q(u_i)|}{|U|} - \frac{1}{|U|}\log_2\frac{|S_Q(u_0)|}{|U|} \\
&= -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_Q(u_i)|}{|U|} = -\sum_{j=1}^{n}\sum_{k=1}^{t_j}\frac{1}{|U|}\log_2\frac{|S_Q(\mu_{jk})|}{|U|} \\
&= -\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\log_2\frac{|Y_j|}{|U|} = H(Q).
\end{aligned}$$

Obviously, $H(Q) < H(P)$. QED

For convenience, the monotonicity of entropy value induced by the partial relation $\preceq'$ is called granulation monotonicity.

Definition 9. Let $S = (U, A)$ be an incomplete information system and $K(A) = (S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|}))$. Information entropy of A is defined as

$$H(A) = -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_A(u_i)|}{|U|}, \tag{8}$$

where $H: A \to [0, \infty)$.

If $K(A)=\omega$, then the information entropy of A achieves its maximum value $\log_2|U|$.

If $K(A)=\delta$, then the information entropy of A achieves its minimum value 0.

Theorem 11. Let $S=(U,A)$ be an incomplete information system and $P,Q\subseteq A$. If $P\prec' Q$, then $H(Q)<H(P)$.

Proof. The proof is similar to that of Theorem 10.

Liang presented a new method for measuring uncertainty in complete information systems, which can be used to measure both uncertainty and fuzziness in rough set theory.

Definition 10[28]. Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. New information entropy of A is defined as

$$E(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{|X_i^c|}{|U|}=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right), \tag{9}$$

where X_i^c denotes the complement set of X_i, i.e., $X_i^c=U-X_i$, $\frac{|X_i|}{|U|}$ represents the probability of X_i within the universe U and $\frac{|X_i^c|}{|U|}$ is the probability of the complement set of X_i within the universe U.

Further on, we gave its condition entropy $E(Q\mid P)=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{|Y_j^c-X_i^c|}{|U|}$ and mutual information $E(P;Q)=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{|Y_j^c\cap X_i^c|}{|U|}$. The relationship between these three concepts is $E(P;Q)=E(Q)-E(Q\mid P)$ [28].

Unlike Shannon's entropy, this entropy can measure not only uncertainty in information systems, but also fuzziness of a rough set and a rough classification.

Theorem 12. Let $S=(U,A)$ be a complete information system and $P,Q\subseteq A$. If $P\prec' Q$, then $E(Q)<E(P)$.

Proof. Similar to Theorem 10, we can obtain that

$$\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right)=\frac{1}{|U|}\left(1-\frac{|S_P(u_{i1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_{i2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_P(u_{is_i})|}{|U|}\right),$$

$$\frac{|Y_j|}{|U|}\left(1-\frac{|Y_j|}{|U|}\right)=\frac{1}{|U|}\left(1-\frac{|S_P(u_{j1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_{j2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_P(u_{jt_j})|}{|U|}\right).$$

Since $P\prec' Q$, one can obtain $S_P(u_i)\subseteq S_Q(u_i)$ and $|S_P(u_i)|\leqslant|S_Q(u_i)|$, $u_i\in U$, and there exists $u_0\in U$ such that $S_P(u_0)\subset S_Q(u_0)$ and $|S_P(u_0)|<|S_Q(u_0)|$. Hence,

$$\begin{aligned}E(P)&=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right)=\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\left(1-\frac{|S_P(u_{ik})|}{|U|}\right)\\&=\sum_{i=1}^{m}\left(\frac{1}{|U|}\left(1-\frac{|S_P(u_{i1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_P(u_{i2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_P(u_{is_i})|}{|U|}\right)\right)\end{aligned}$$

$$= \sum_{i=1}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_P(u_i)|}{|U|}\right) = \sum_{i=1, u_i \neq u_0}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_P(u_i)|}{|U|}\right) + \frac{1}{|U|}\left(1 - \frac{|S_P(u_0)|}{|U|}\right)$$

$$> \sum_{i=1, u_i \neq u_0}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_Q(u_i)|}{|U|}\right) + \frac{1}{|U|}\left(1 - \frac{|S_Q(u_0)|}{|U|}\right)$$

$$= \sum_{i=1}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_Q(u_i)|}{|U|}\right) = \sum_{i=1}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_Q(u_i)|}{|U|}\right) = \sum_{j=1}^{n}\sum_{k=1}^{t_j} \frac{1}{|U|}\left(1 - \frac{|S_Q(u_{jk})|}{|U|}\right)$$

$$= \sum_{j=1}^{n} \frac{|Y_j|}{|U|}\left(1 - \frac{|Y_j|}{|U|}\right) = E(Q).$$

Obviously, $E(Q) < E(P)$.

Definition 11. Let $S = (U, A)$ be an incomplete information system and $K(A) = (S_A(u_1), S_A(u_2), \cdots, S_A(u_{|U|}))$. Information entropy of A is defined as

$$E(A) = \sum_{i=1}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_A(u_i)|}{|U|}\right). \tag{10}$$

If $K(A) = \omega$, then the information entropy of A achieves its maximum value $1 - 1/|U|$.

If $K(A) = \delta$, then the information entropy of A achieves its minimum value 0.

Theorem 13. Let $S = (U, A)$ be a complete information system, $U/IND(A) = \{X_1, X_2, \cdots, X_m\}$ and $K(A) = (S_A(u_1), S_A(u_2), \cdots, S_A(u_{|U|}))$. Information entropy of A degenerates into

$$E(A) = \sum_{i=1}^{m} \frac{|X_i|}{|U|}\left(1 - \frac{|X_i|}{|U|}\right),$$

i.e.,

$$E(A) = \sum_{i=1}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_A(u_i)|}{|U|}\right) = \sum_{i=1}^{m} \frac{|X_i|}{|U|}\left(1 - \frac{|X_i|}{|U|}\right).$$

Proof. The proof is similar to that of Theorem 12.

Theorem 13 shows that the information entropy in complete information systems is a special form of the information entropy in incomplete information systems.

Theorem 14. Let $S = (U, A)$ be an incomplete information system and $P, Q \subseteq A$. If $P \prec' Q$, then $E(Q) < E(P)$.

Proof. The proof is similar to that of Theorem 12.

Definition 12[21]. Let $S = (U, A)$ be an incomplete information system and $K(A) = (S_A(u_1), S_A(u_2), \cdots, S_A(u_{|U|}))$. Combination entropy of A is defined as

$$CE(A) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{C_{|U|}^2 - C_{|S_A(u_i)|}^2}{C_{|U|}^2}, \tag{11}$$

where $\frac{C_{|U|}^2 - C_{|S_A(u_i)|}^2}{C_{|U|}^2}$ denotes the probability of pairs of the elements which are probably distinguishable to each other within the whole number of pairs of the elements on the universe U.

If $K(A)=\omega$, then the combination entropy of A achieves its maximum value 1.

If $K(A)=\delta$, then the combination entropy of A achieves its minimum value 0.

Theorem 15. Let $S=(U,A)$ be an incomplete information system and $P,Q\subseteq A$. If $P\prec' Q$, then $CE(Q)<CE(P)$.

Proof. Suppose $K(P)=(S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|}))$ and $K(Q)=(S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|}))$. Since $P\prec' Q$, we have that $S_P(u_i)\subseteq S_Q(u_i)$ and $|S_P(u_i)|\leqslant|S_Q(u_i)|$, $u_i\in U$, and there exists $u_0\in U$ such that $S_P(u_0)\subset S_Q(u_0)$ and $|S_P(u_0)|<|S_Q(u_0)|$.

Therefore,

$$CE(P)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2-C_{|S_P(u_i)|}^2}{C_{|U|}^2}=1-\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}=1-\frac{1}{|U|}\sum_{i=1,u_i\neq u_0}^{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}-\frac{1}{|U|}\frac{C_{|S_P(u_0)|}^2}{C_{|U|}^2}$$

$$>1-\frac{1}{|U|}\sum_{i=1,u_i\neq u_0}^{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}-\frac{1}{|U|}\frac{C_{|S_Q(u_0)|}^2}{C_{|U|}^2}=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2-C_{|S_Q(u_i)|}^2}{C_{|U|}^2}=CE(Q).$$

Obviously, $CE(Q)<CE(P)$.

Theorem 16. Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. Then, the combination entropy of A degenerates into

$$CE(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|U|}^2-C_{|X_i|}^2}{C_{|U|}^2}, \tag{12}$$

where $\frac{C_{|U|}^2-C_{|X_i|}^2}{C_{|U|}^2}$ denotes the probability of pairs of the elements which are distinguishable each other within the whole number of pairs of the elements on the universe U.

Proof. Suppose $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$, $X_i=\{u_{i1},u_{i2},\cdots,u_{is_i}\}$, $i\in\{1,2,\cdots,m\}$, where $||X_i|=s_i$ and $\sum_{i=1}^{m}s_i=|U|$. Let $K(A)=(S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|}))$, then the relationship between $K(A)$ and $U/IND(A)$ is as follows:

$$X_i=S_P(u_{i1})=S_P(u_{i2})=\cdots=S_P(u_{is_i}),$$

i.e.,

$$|X_i|=|S_P(u_{i1})|=|S_P(u_{i2})|=\cdots=|S_P(u_{is_i})|.$$

Hence, one can obtain that

$$|X_i|\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=s_i\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=\sum_{k=1}^{s_i}\frac{C_{|S_A(u_{ik})|}^2}{C_{|U|}^2}.$$

Therefore,

$$CE(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C_{|U|}^2-C_{|X_i|}^2}{C_{|U|}^2}=1-\frac{1}{|U|}\sum_{i=1}^{m}|X_i|\times\frac{C_{|X_i|}^2}{C_{|U|}^2}=1-\frac{1}{|U|}\sum_{i=1}^{m}s_i\times\frac{C_{|X_i|}^2}{C_{|U|}^2}$$

$$=1-\frac{1}{|U|}\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{C^2_{|S_A(u_{ik})|}}{C^2_{|U|}}=1-\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|S_A(u_i)|}}{C^2_{|U|}}=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C^2_{|U|}-C^2_{|S_A(u_i)|}}{C^2_{|U|}}.$$

This completes the proof.

5 The relationship between information granulation and entropy

In some sense, there exists a complement relation between entropy and information granulation. That is to say, the bigger the entropy value is, the smaller the information granulation; the smaller the entropy value is, the bigger the information granulation.

5.1 The relationship between information granulation and entropy in complete information systems

Theorem 17[18]**.** Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. Then, the relationship between the information entropy $E(A)$ and the information granulation $GK(A)$ is as follows:

$$E(A)+GK(A)=1. \tag{13}$$

Theorem 18[18]**.** Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. Then, the relationship between Shannon' entropy $H(A)$ and the rough entropy $E_r(A)$ is as follows:

$$H(A)+E_r(A)=\log_2|U|. \tag{14}$$

Theorem 19. Let $S=(U,A)$ be a complete information system and $U/IND(A)=\{X_1,X_2,\cdots,X_m\}$. Then, the relationship between the rough entropy $CE(A)$ and the combination granulation $CG(A)$ is as follows:

$$CE(A)+CG(A)=1. \tag{15}$$

Proof. From their definitions, one can obtain that

$$CE(A)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C^2_{|U|}-C^2_{|X_i|}}{C^2_{|U|}}=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{C^2_{|X_i|}}{C^2_{|U|}}\right)=1-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}}=1-CG(A).$$

Clearly, $CE(A)+CG(A)=1$.

5.2 The relationship between information granulation and entropy in incomplete information systems

Theorem 20. Let $S=(U,A)$ be an incomplete information system and $K(A)=(S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|}))$. Then, the relationship between the information entropy $E(A)$ and the information granulation $GK(A)$ is as follows:

$$E(A)+GK(A)=1. \tag{16}$$

Proof. Let $K(A)=(S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|}))$, then

$$E(A)=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_A(u_i)|}{|U|}\right)=\sum_{i=1}^{|U|}\frac{1}{|U|}-\sum_{i=1}^{|U|}\frac{|S_A(u_i)|}{|U|^2}=1-GK(A).$$

i.e., $E(A)+GK(A)=1$.

Theorem 21. Let $S=(U,A)$ be an incomplete information system and $K(A)=(S_A(u_1), S_A(u_2),\cdots,S_A(u_{|U|}))$. Then, the relationship between the information entropy $H(A)$ and the rough entropy $E_r(A)$ is as follows:

$$H(A)+E_r(A)=\log_2|U|. \tag{17}$$

Proof. It follows from Definition 7 and Definition 9 that

$$H(A)=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_A(u_i)|}{|U|}=-\sum_{i=1}^{|U|}\frac{1}{|U|}(\log_2|S_A(u_i)|-\log_2|U|)$$

$$=-\left(-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_A(u_i)|}\right)+\log_2|U|\sum_{i=1}^{|U|}\frac{1}{|U|}=-E_r(A)+\log_2|U|.$$

That is $H(A)+E_r(A)=\log_2|U|$.

Theorem 22. Let $S=(U,A)$ be an incomplete information system and $K(A)=(S_A(u_1), S_A(u_2),\cdots,S_A(u_{|U|}))$. Then, the combination entropy $CE(A)$ and the combination granulation $CG(A)$ is as follows:

$$CE(A)+CG(A)=1. \tag{18}$$

Proof. From their definitions, one know that

$$CE(A)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|U|}^2-C_{|S_A(u_i)|}^2}{C_{|U|}^2}=1-\sum_{i=1}^{|U|}\frac{C_{|S_A(u_i)|}^2}{C_{|U|}^2}=1-CG(A).$$

Obviously, $CE(A)+CG(A)=1$.

6 Conclusions

Information granulation and entropy theory are two main approaches to research the uncertainty of an information system, which have been widely applied in many practical issues. In this paper, the characterizations and representations of information granules in complete information systems and incomplete information systems have been analyzed and an axiom approach to information granulation has been presented. It has been proved that each of the existing information granulations is a special instance of this axiom definition. Furthermore, some results of entropy theory have been extended in information systems and the relationship between information granulation and entropy has been established. It deserves to be pointed out that entropy measures in information systems all satisfy granulation monotonicity. These results unify the relative conclusions about uncertainty measure in complete information systems and incomplete information systems. Further research work is planned to establish an axiom approach of fuzzy information granulation in fuzzy information systems.

1 Zadeh L A. Fuzzy logic=computing with words. IEEE Trans Fuzzy Syst, 1996, 4(1): 103－111

2 Zadeh L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst, 1997, 19(1): 111－127

3 Zadeh L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent system. Soft Comp, 1998, 2(1): 23－25

4 Zadeh L A. Fuzzy sets and information granularity. In: Gupta N, Ragade R, Yager R, eds. Advances in Fuzzy Set Theory and Application. Amsterdam: North-Holland, 1979. 3－18

5 Zhang W X, Wei L, Qi J J. Theory and method of attribute reduction in concept lattice. Sci China Ser E (in Chinese), 2005, 35(6): 628－639

6 Zhang W X, Qiu G F, Wu W Z. Attribute reduction theory and approach to concept lattice. Sci China Ser F-Inf Sci, 2005, 48(6): 713－726

7 Pawlak Z. Granularity of knowledge, indiscernibility and rough sets. In: Proceedings of 1998 IEEE International Conference on Fuzzy Systems, Heideberg: Physifca-Verlag, 1998. 106－110

8 Lin T Y. Granular computing on binary relations I: data mining and neighborhood systems, II: rough sets representations and belief functions. In: Polkowski L, Skowron A, eds. Rough Sets in Knowledge Discovery 1. Heidelberg: Physica-Verlag, 1998. 107－140

9 Klir G J. Basic issues of computing with granular computing. In: Proceedings of 1998 IEEE International Conference on Fuzzy Systems. 1998. 101－105

10 Liang J Y, Qian Y H, Chu C Y, et al. Rough set approximation based on dynamic granulation. Lecture Notes in Artificial Intelligence, 2005, 3641: 701－708

11 Zhang J S, Leung Y, Xu Z B. Clustering methods by simulating visual systems. J Comp (in Chinese), 2001, 24(5): 496－501

12 Shi Z Z. Advanced Artificial Intelligence. Beijing: Science Press, 2006

13 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic Publishers, 1991

14 Zhang B, Zhang L. Theory of Problem Solving with Applications. Beijing: Tsinghua University Press, 1990

15 Zhang L, Zhang B. Theory of fuzzy quotient space (methods of fuzzy granular computing). J Software (in Chinese), 2003, 14(4): 770－776

16 Zhang L, Zhang B. Fuzzy reasoning model under quotient space structure. Inf Sci, 2005, 173: 353－364

17 Wierman M J. Measuring uncertainty in rough set theory. International J General Sys, 1999, 28(4): 283－297

18 Liang J Y, Shi Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory. Int J Uncertainty, Fuzziness and Knowledge-Based Systems, 2004, 12(1): 37－46

19 Liang J Y, Xu Z B. The algorithm on knowledge reduction in incomplete information systems. Int J Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 24(1): 95－103

20 Liang J Y, Li D Y. Uncertainty and Knowledge Acquisition in Information Systems. Beijing: Science Press, 2005

21 Qian Y H, Liang J Y. Combination entropy and combination granulation in incomplete information system. Lecture Notes in Artificial Intelligence, 2006, 4062: 184－190

22 Shannon C E. The mathematical theory of communication. Bell Sys Tech J, 1948, 27(3, 4): 373－423, 623－656

23 Düntsch I, Gediga G. Uncertainty measures of rough set prediction. Artif Intell, 1998, 106: 283－297

24 Beaubouef T, Petry F E. Fuzzy rough set techniques for uncertainty processing in a relational database. Int J Intell Syst, 2000, 15(5): 389－424

25 Beaubouef T, Petry F E, Arora G. Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inf Sci, 1998, 109(1-4): 185－195

26 De Luca A, Termini S. A definition of nonprobilistic entropy in the setting of fuzzy sets theory. Inf Contr, 1972, 20: 301－312

27 Mi J S, Leung Y, Wu W Z. An uncertainty measure in partition-based fuzzy rough sets. Int J General Syst, 2005, 34(1): 77－90

28 Liang J Y, Chin K S, Dang C Y, et al. A new method for measuring uncertainty and fuzziness in rough set theory. Int J General Syst, 2002, 31(4): 331－342

29 Kryszkiewicz M. Rough set approach to incomplete information systems. Inf Sci, 1998, 112: 39－49