International Journal of General Systems Vol. 35 No. 6, December, 2006, 641–654

Taylor & Francis
Taylor & Francis Group

Information entropy, rough entropy and knowledge granulation in incomplete information systems

J. LIANG†‡*, Z. SHI‡, D. LI† and M. J. WIERMAN¶

†Key Laboratory of Ministry of Education for Computation Intelligence and Chinese Information Processing, School of Computer and Information Technology, Shanxi University, Taiyuan 030006, People's Republic of China

‡Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, People's Republic of China

¶Creighton University, Omaha, NE 68005, USA

(Received 15 April 2005; in final form 19 January 2006)

Rough set theory is a relatively new mathematical tool for use in computer applications in circumstances that are characterized by vagueness and uncertainty. Rough set theory uses a table called an information system, and knowledge is defined as classifications of an information system. In this paper, we introduce the concepts of information entropy, rough entropy, knowledge granulation and granularity measure in incomplete information systems, their important properties are given, and the relationships among these concepts are established. The relationship between the information entropy $E(A)$ and the knowledge granulation $GK(A)$ of knowledge A can be expressed as $E(A) + GK(A) = 1$, the relationship between the granularity measure $G(A)$ and the rough entropy $E_r(A)$ of knowledge A can be expressed as $G(A) + E_r(A) = \log_2|U|$. The conclusions in Liang and Shi (2004) are special instances in this paper. Furthermore, two inequalities $-\log_2 GK(A) \leq G(A)$ and $E_r(A) \leq \log_2(|U|(1 - E(A)))$ about the measures GK, G, E and E_r are obtained. These results will be very helpful for understanding the essence of uncertainty measurement, the significance of an attribute, constructing the heuristic function in a heuristic reduct algorithm and measuring the quality of a decision rule in incomplete information systems.

Keywords: Incomplete information systems; Rough sets; Information entropy; Rough entropy; Knowledge granulation

1. Introduction

Rough set theory, introduced by Pawlak (1991) and Pawlak *et al.* (1995), is a relatively new soft computing tool for the analysis of a vague description of an object. The adjective "vague", referring to the quality of information, means inconsistency or ambiguity which follows from information granulation. The rough set philosophy is based on the assumption that with every object of the universe there is associated a certain amount of information (data, knowledge), expressed by means of some attributes used for object description. Objects having the same description are indiscernible (similar) with respect to the available information. The indiscernibility relation thus generated constitutes a mathematical basis of the rough set theory; it induces a partition of the universe into blocks of indiscernible objects,

*Corresponding author. Email: ljy@sxu.edu.cn

International Journal of General Systems
ISSN 0308-1079 print/ISSN 1563-5104 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/03081070600687668

called elementary sets, that can be used to build knowledge about a real or abstract world (Pawlak 1991, 1995, Zhang *et al.* 2001, Leung and Li 2003, Mi *et al.* 2004). The use of the indiscernibility relation results in information granulation.

The entropy of a system, as defined by Shannon (1948) gives a measure of uncertainty about its actual structure. It has been a useful mechanism for characterizing the information content in various modes and applications in many diverse fields. Several authors (Beaubouef *et al.* 1998, Düntsch and Gediga 1998, Klir and Wierman 1999, Chakik *et al.* 2004) have used Shannon's entropy and its variants to measure uncertainty in rough set theory. A new definition for information entropy in rough set theory is presented by Liang *et al.* (2002). Unlike the logarithmic behaviour of Shannon entropy, the gain function considered there possesses the complement nature. Wierman (1999) presented a well-justified measure of uncertainty, the measure of granularity, along with an axiomatic derivation. Its strong connections to the Shannon entropy and the Hartley measure of uncertainty also lend strong support to its correctness and applicability. Furthermore, the relationships among information entropy, rough entropy and knowledge granulation in complete information systems have been established by Liang and Shi (2004) and Liang and Li (2005).

In the paper, the concepts of information entropy, rough entropy, knowledge granulation and granularity measure in incomplete information systems are introduced. The relationships among these concepts are established. The conclusions in Liang and Shi (2004) are generalized. These results will be very helpful for understanding the essence of uncertainty measurement, the significance of an attribute, constructing the heuristic function in a heuristic reduct algorithm and measuring the quality of a decision rule in incomplete information systems.

2. Incomplete information system

An information system is a pair $S = (U, A)$, where

(i) U is a non-empty finite set of objects;
(ii) A is a non-empty finite set of attributes;
(iii) for every $a \in A$, there is a mapping a, a: $U \to V_a$, where V_a is called the value set of a.

Each subset of attributes $P \subseteq A$ determines a binary indistinguishable relation IND(P) as follows

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

It can be easily shown that IND(P) is an equivalence relation on the set U.

For $P \subseteq A$, the relation IND(P) constitutes a partition of U, which is denoted by U/IND(P).

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.

If V_a contains a null value for at least one attribute $a \in A$, then S is called an incomplete information system (Kryszkiewicz 1998, 1999), otherwise it is complete. Further on, we will denote the null value by $*$.

Let $S = (U, A)$ be an information system, $P \subseteq A$ an attribute set. We define a binary relation on U as follows

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

In fact, SIM(P) is a tolerance relation on U (Kryszkiewicz 1998), the concept of a tolerance relation has a wide variety of applications in classification (Kryszkiewicz 1998).

It can be easily shown that $\mathrm{SIM}(P) = \cap_{a \in P} \mathrm{SIM}(\{a\})$.

Let $S_P(u)$ denote the set $\{v \in U | (u, v) \in \mathrm{SIM}(P)\}$. $S_P(u)$ is the maximal set of objects which are possibly indistinguishable by P with u.

Let $U/\mathrm{SIM}(P)$ denote the family sets $\{S_P(u) | u \in U\}$, the classification induced by P. A member $S_P(u)$ from $U/\mathrm{SIM}(P)$ will be called a tolerance class or a granule of information. It should be noticed that the tolerance classes in $U/\mathrm{SIM}(P)$ do not constitute a partition of U in general. They constitute a covering of U, i.e. $S_P(u) \neq \emptyset$ for every $u \in U$, and $\cup_{u \in U} S_P(u) = U$.

Let $S = (U, A)$ be an information system, $X \subseteq U$ and $P \subseteq A$. $\underline{P}X$ is the lower approximation to X, if

$$\underline{P}X = \{x \in U | S_P(x) \subseteq X\} = \{x \in X | S_P(x) \subseteq X\}.$$

$\bar{P}X$ is the upper approximation to X, if

$$\bar{P}X = \{x \in U | S_P(x) \cap X \neq \emptyset\} = \cup\{S_P(x) | x \in X\}.$$

As in complete information systems, $\underline{P}X$ is a set of objects that belong to X with certainty, and $\bar{P}X$ is a set of objects that possibly belong to X.

Example 1. Consider descriptions of several cars in table 1 (Liang and Xu 2002).

This is an incomplete information system, where $U = \{u_1, u_2, u_3, u_4, u_5\}$, and $A = \{a_1, a_2, a_3, a_4\}$ with a_1-price, a_2-size, a_3-engine, a_4-max-speed. By computing, it follows that $U/\mathrm{SIM}(A) = \{S_A(u_1), S_A(u_2), S_A(u_3), S_A(u_4), S_A(u_5)\}$, where $S_A(u_1) = \{u_1\}$, $S_A(u_2) = \{u_2\}$, $S_A(u_3) = S_A\{u_4\} = \{u_3, u_4\}$, $S_A(u_5) = \{u_5\}$.

Now we define a partial order on the set of all classifications of U. Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$. We say that Q is coarser than P (or P is finer than Q), denoted by $P \preccurlyeq Q$, if and only if $S_P(u_i) \subseteq S_Q(u_i)$ for $\forall i \in \{1, 2, \ldots, |U|\}$. If $P \preccurlyeq Q$ and $P \neq Q$, we say that Q is strictly coarser than P (or P is strictly finer than Q) and denoted by $P \prec Q$.

In fact, $P \prec Q \Leftrightarrow$ for $\forall i \in \{1, 2, \ldots, |U|\}$, we have that $S_P(u_i) \subseteq S_Q(u_i)$, and $\exists j \in \{1, 2, \ldots, |U|\}$, such that $S_P(u_j) \subset S_Q(u_j)$.

3. Information entropy and knowledge granulation

In this section, information entropy and knowledge granulation in an incomplete information system are introduced. The properties of these are discussed respectively, and the relationship between them is established.

Table 1. The information system about the car.

Car	*Price*	*Size*	*Engine*	*Max-speed*
u_1	Low	Compact	*	Low
u_2	Low	Full	Diesel	High
u_3	High	Full	Diesel	Medium
u_4	High	*	Diesel	Medium
u_5	Low	Full	Gasoline	High

Let $S = (U, A)$ be an information system. By $U/\mathrm{SIM}(A) = \{S_A(u) | u \in U\}$, we denote the classification induced by A. Of particular interest is the discrete classification

$$\hat{A}(U) = \{S_A(u) = \{u\} | u \in U\} \tag{1}$$

and the indiscrete classification

$$\check{A}(U) = \{S_A(u) = U | u \in U\} \tag{2}$$

or just $\hat{A}$ and $\check{A}$ if there is no confusion as to the domain set involved.

Definition 1. Let $S = (U, A)$ be an incomplete information system. The information entropy of knowledge A is defined as

$$E(A) = \sum_{i=1}^{|U|} \frac{1}{|U|}\left(1 - \frac{|S_A(u_i)|}{|U|}\right) \tag{3}$$

where

$$1 - \frac{|S_A(u_i)|}{|U|}$$

represents the probability of the complement of $S_A(u_i)$ within the universe U.

If $U/\mathrm{SIM}(A) = \hat{A}$, then the information entropy of knowledge A achieves the maximum value $1 - 1/|U|$.

If $U/\mathrm{SIM}(A) = \check{A}$, then the information entropy of knowledge A achieves the minimum value 0.

Obviously, for an incomplete information system $S = (U, A)$, we have that $0 \leq E(A) \leq 1 - 1/|U|$.

Proposition 1. Let $S = (U, A)$ be a complete information system, $U/\mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$. Then the information entropy of knowledge A degenerates into

$$E(A) = \sum_{i=1}^{m} \frac{|X_i|}{|U|}\left(1 - \frac{|X_i|}{|U|}\right). \tag{4}$$

Proof. Suppose that $U/\mathrm{SIM}(A) = \{S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|})\}$ and $X_i = \{u_{i1}, u_{i2}, \ldots, u_{is_i}\}$ $(i = 1, 2, \ldots, m)$. It follows that $|X_i| = s_i$ and $\sum_{i=1}^{m} |s_i| = |U|$.

It is easy to know that $X_i = S_A(u_{i1}) = S_A(u_{i2}) = \cdots = S_A(u_{is_i})$ and $|X_i| = |S_A(u_{i1})| = |S_A(u_{i2})| = \cdots = |S_A(u_{is_i})|$ for $i = 1, 2, \ldots, m$. Then, we have that

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

Hence

$$\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left(1-\frac{|X_i|}{|U|}\right)=\sum_{i=1}^{m}\left(\frac{1}{|U|}\left(1-\frac{|S_A(u_{i1})|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_A(u_{i2})|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_A(u_{is_i})|}{|U|}\right)\right)$$

$$=\left(\frac{1}{|U|}\left(1-\frac{|S_A(u_1)|}{|U|}\right)+\frac{1}{|U|}\left(1-\frac{|S_A(u_2)|}{|U|}\right)+\cdots+\frac{1}{|U|}\left(1-\frac{|S_A(u_{|U|})|}{|U|}\right)\right)$$

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

This completes the proof. □

In Liang *et al.* (2002), the information entropy of a complete information system $S=(U,A)$ with $U/\mathrm{IND}(A)=\{X_1, X_2, \ldots, X_m\}$ is defined as

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

Proposition (1) states that the information entropy in complete information systems is a special instance of that in incomplete information systems. It means that the definition of information entropy of incomplete information systems is a consistent extension to that of complete information systems.

PROPOSITION 2. Let $S=(U, A)$ be an incomplete information system, $P, Q\subseteq A$. If $P\prec Q$, then $E(Q)<(E(P)$.

Proof. From $P\prec Q$, it follows that $S_P(u_i)\subseteq S_Q(u_i)$ $(\forall i\in\{1,2,\ldots,|U|\})$, and $\exists j\in\{1,2,\ldots,|U|\}$ such that $S_P(u_j)\subset S_Q(u_j)$. By $|S_P(u_j)|<(|S_Q(u_j)|$, we have that

$$E(Q)=\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_Q(u_i)|}{|U|}\right)<\sum_{i=1}^{|U|}\frac{1}{|U|}\left(1-\frac{|S_P(u_i)|}{|U|}\right)=E(P).$$

This completes the proof. □

Proposition (2) states that the information entropy of knowledge increases as tolerance classes become smaller through finer classification.

DEFINITION 2. Let $S = (U, A)$ be an incomplete information system. The granulation of knowledge A is defined as

$$GK(A) = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_A(u_i)| = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_A(u_i)|}{|U|} \tag{5}$$

where $|S_A(u_i)|/|U|$ represents the probability of tolerance class $S_A(u_i)$ within the universe U.

If $U/\mathrm{SIM}(A) = \hat{A}$, then the granulation of knowledge A achieves the minimum value $|U|/|U|^2 = 1/|U|$.

If $U/\mathrm{SIM}(A) = \check{A}$, then the granulation of knowledge A achieves the maximum value $|U|^2/|U|^2 = 1$.

Obviously, for an incomplete information system $S = (U, A)$, we have that $1/|U| \leq GK(A) \leq 1$. Knowledge granulation can represent the discernibility ability of knowledge, the smaller $GK(A)$ is, the stronger its discernibility ability is.

PROPOSITION 3. Let $S = (U, A)$ be a complete information system, $U/\mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$. Then the granulation of knowledge A degenerates into

$$GK(A) = \frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2. \tag{6}$$

Proof. Suppose that $U/\mathrm{SIM}(A) = \{S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|})\}$ and $X_i = \{u_{i1}, u_{i2}, \ldots, u_{is_i)}\}$. It follows that $|X_i| = s_i$ and $|X_i| = s_i$ and $\sum_{i=1}^{m}|s_i| = |U|$.

It is easy to show that $X_i = S_A(u_{i1}) = S_A(u_{i2}) = \cdots = S_A(u_{is_i})$ and $|X_i| = |S_A(u_{i1})| = |S_A(u_{i2})| = \cdots = |S_A(u_{is_i})|$ for $i = 1, 2, \ldots, m$.

Noticing that

$$|X_i|^2 = S_i^2 = |S_A(u_{i1})| + |S_A(u_{i2})| + \cdots + |S_A(u_{is_i})|$$

therefore

$$\begin{aligned}
\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2 &= \frac{1}{|U|^2}\sum_{i=1}^{m}(|S_A(u_{i1})| + |S_A(u_{i2})| + \cdots + |S_A(u_{is_i})|) \\
&= \frac{1}{|U|^2}(|S_A(u_1)| + |S_A(u_2)| + \cdots + |S_A(u_{|U|})|) \\
&= \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_A(u_i)| = GK(A).
\end{aligned}$$

This completes the proof. □

In Liang and Shi (2004), the knowledge granulation in a complete information system $S = (U, A)$ with $U/\mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$ is defined as

$$\frac{1}{|U|^2}\sum_{i=1}^{m}|X_i|^2$$

Proposition (3) states that the knowledge granulation in complete information systems is a special instance of that in incomplete information systems. It means that the definition of knowledge granulation in incomplete information systems is a consistent extension to that of complete information systems.

PROPOSITION 4. Let $S = (U, A)$ be an incomplete information system, P, $Q \subseteq A$. If $P \prec Q$, then $GK(P) < (GK(Q)$.

Proof. From $P \prec Q$, it follows that $S_P(u_i) \subseteq S_Q(u_i)$ $(\forall i \in \{1, 2, \ldots, |U|\})$, and $\exists j \in \{1, 2, \ldots, |U|\}$ such that $S_P(u_j) \subset S_Q(u_j)$. By $|S_P(u_j)| < (|S_Q(u_j)|$, we have that

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

This completes the proof. □

Proposition (4) states that knowledge granulation decreases as tolerance classes become smaller through finer classification.

PROPOSITION 5. For arbitrary incomplete information system $S = (U, A)$, we have that

$$E(A) + GK(A) = 1. \tag{7}$$

Proof. Let $S = (U, A)$ be an incomplete information system, $U/\mathrm{SIM}(A) = \{S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|})\}$. By the definitions 1 and 2,

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

It follows that $E(A) + GK(A) = 1$. This completes the proof. □

Proposition (5) shows that information entropy $E(A)$ and knowledge granulation $GK(A)$ possess the same capability on depicting the uncertainty of an information system.

Example 2. For table I, $U = \{u_1, u_2, u_3, u_4, u_5\}$, $A = \{\text{price, size, engine, max-speed}\}$, $U/\mathrm{SIM}(A) = \{\{u_1\}, \{u_2\}, \{u_3, u_4\}, \{u_3, u_4\}, \{u_5\}\}$. By computing, it follows that

$$\begin{aligned} E(A) &= \sum_{i=1}^{|U|}\frac{1}{|U|}\left(1 - \frac{|S_A(u_i)|}{|U|}\right) \\ &= \frac{1}{5}\left[\left(1-\frac{1}{5}\right) + \left(1-\frac{1}{5}\right) + \left(1-\frac{2}{5}\right) + \left(1-\frac{2}{5}\right) + \left(1-\frac{1}{5}\right)\right] = \frac{18}{25} \end{aligned}$$

and

$$GK(A) = \frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_A(u_i)| = \frac{1}{25}(1+1+2+2+1) = \frac{7}{25}.$$

It is clear that $E(A) + GK(A) = 1$, i.e. the sum of the information entropy and the granulation of knowledge A is constant 1.

4. Granularity measure and rough entropy

In this section, the granularity measure and the rough entropy of an incomplete information system are introduced. The properties of them are discussed respectively. The relationship between them is established.

DEFINITION 3. Let $S=(U,A)$ be an incomplete information system. The granulation measure of S is defined as

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

where $G{:}R\longrightarrow[0,\infty)$ is a function from the set R of all classifications $U/\mathrm{SIM}(A)$ to the set of non-negative real numbers and $(|S_A(u_i)|)/|U|$ denotes the probability of tolerance class $S_A(u_i)$ within the universe U.

The granularity measure, $G(A)$, measures the uncertainty associated with the prediction of outcomes where elements of each class $S_A(u_i)$ of $U/\mathrm{SIM}(A)$ are indistinguishable.

If $U/\mathrm{SIM}(A)=\hat{A}$, then the granularity measure $G(A)$ achieves the maximum value $\log_2|U|$.

If $U/\mathrm{SIM}(A)=\check{A}$, then the granularity measure $G(A)$ achieves the minimum value 0.

Obviously, for an information system $S=(U,A)$, we have that $0\leq G(A)\leq\log_2|U|$.

PROPOSITION 6. Let $S=(U,A)$ be a complete information system and $U/\mathrm{IND}(A)=\{X_1, X_2,\ldots,X_m\}$. Then the granularity measure of S degenerates into

$$G(A)=-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}. \tag{9}$$

Proof. Suppose that $U/\mathrm{SIM}(A)=\{S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|})\}$ and $X_i=\{u_{i1},u_{i2},\ldots,u_{is_i}\}$ $(i=1,2,\ldots,m)$. It follows that $|X_i|=s_i$ and $\sum_{i=1}^{m}|s_i|=|U|$.

It is easy to know that $X_i=S_A(u_{i1})=S_A(u_{i2})=\cdots=S_A(u_{is_i})$ and $|X_i|=|S_A(u_{i1})|=|S_A(u_{i2})|=\cdots=|S_A(u_{is_i})|$ for $i=1,2,\ldots,m$.

Then, we have that

$$\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|}=\frac{1}{|U|}\log_2\frac{|S_A(u_{i1})|}{|U|}+\frac{1}{|U|}\log_2\frac{|S_A(u_{i2})|}{|U|}+\cdots+\frac{1}{|U|}\log_2\frac{|S_A(u_{is_i})|}{|U|}.$$

Hence

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

This completes the proof. □

In Chakik *et al.* (2004) and Wierman (1999), the granularity measure of a complete information system $S = (U, A)$ with $U/\mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$ is defined as $-\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|}$. Proposition (6) states that the granularity measure of complete information systems is a special instance of incomplete information systems. It means that the definition of the granularity measure of incomplete information systems is a consistent extension to that of complete information systems.

PROPOSITION 7. Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$. If $P \prec Q$, then $G(Q) < (G(P)$.

Proof. From $P \prec Q$, it follows that $S_P(u_i) \subseteq S_Q(u_i)$ $(\forall i \in \{1, 2, \ldots, |U|\})$, and $\exists j \in \{1, 2, \ldots, |U|\}$ such that $S_P(u_j) \subset S_Q(u_j)$. By $|S_P(u_j)| < (|S_Q(u_j)|$, we have that

$$G(Q) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_Q(u_i)|}{|U|} < -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_P(u_i)|}{|U|} = G(P).$$

This completes the proof. □

Proposition (7) states that the granularity measure increases as tolerance classes become smaller through finer classification.

The concept of rough entropy has been introduced in rough sets, rough relational databases and information systems (Beaubouef *et al.* 1998, Liang and Shi 2004). Now we introduce a definition of rough entropy of knowledge in incomplete information systems.

DEFINITION 4. Let $S = (U, A)$ be an incomplete information system, the rough entropy of knowledge A is defined as

$$E_{\mathrm{r}}(A) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_i)|} \tag{10}$$

where $1/(|S_A(u_i)|)$ represents the probability of an element within the tolerance class $S_A(u_i)$.

If $U/\mathrm{SIM}(A) = \hat{A}$, then the rough entropy of knowledge A achieves the minimum value 0.

If $U/\mathrm{SIM}(A) = \check{A}$, then the rough entropy of knowledge A achieves the maximum value $\log_2|U|$.

Obviously, for an information system $S = (U, A)$, we have that $0 \leq E_{\mathrm{r}}(A) \leq \log_2|U|$.

PROPOSITION 8. Let $S = (U, A)$ be a complete information system and $U/\mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$. Then the rough entropy of knowledge A degenerates into

$$E_{\mathrm{r}}(A) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \log_2 \frac{1}{|X_i|}. \tag{11}$$

Proof. Suppose that $U/\mathrm{SIM}(A) = \{S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|})\}$ and $X_i = \{u_{i1}, u_{i2}, \ldots, u_{is_i}\}$ $(i = 1, 2, \ldots, m)$. It follows that $|X_i| = s_i$ and $\sum_{i=1}^{m} |s_i| = |U|$.

It is easy to know that $X_i = S_A(u_{i1}) = S_A(u_{i2}) = \cdots = S_A(u_{is_i})$ and $|X_i| = |S_A(u_{i1})| = |S_A(u_{i2})| = \cdots = |S_A(u_{is_i})|$ for $i = 1, 2, \ldots, m$.

Then, we have that

$$\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|} = \frac{1}{|U|}\log_2\frac{1}{|S_A(u_{i1})|} + \frac{1}{|U|}\log_2\frac{1}{|S_A(u_{i2})|} + \cdots + \frac{1}{|U|}\log_2\frac{1}{|S_A(u_{is_i})|}.$$

Hence

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

This completes the proof. □

In Liang and Shi (2004), the rough entropy of knowledge in a complete information system $S = (U, A)$ with $U/\mathrm{IND}(A) = \{X_1, X_2, \ldots, X_m\}$ is defined as

$$-\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|}$$

Proposition (8) states that the rough entropy in complete information systems is a special instance of that in incomplete information systems. It means that the definition of the rough entropy in incomplete information systems is a consistent extension to that in complete information systems.

Proposition 9. Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$. If $P \prec Q$, then $E_{\mathrm{r}}(P) < E_{\mathrm{r}}(Q)$.

Proof. From $P \prec Q$, it follows that $S_P(u_i) \subseteq S_Q(u_i) \quad (\forall i \in \{1, 2, \ldots, |U|\})$, and $\exists j \in \{1, 2, \ldots, |U|\}$ such that $S_P(u_j) \subset S_Q(u_j)$. By $|S_P(u_j)| < |S_Q(u_j)|$, we have that

$$\begin{aligned}
E_{\mathrm{r}}(P) &= -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)|}\\
&< -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_Q(u_i)|}\\
&= E_{\mathrm{r}}(Q)
\end{aligned}$$

This completes the proof. □

Proposition (9) states that the rough entropy of knowledge decreases as tolerance classes become smaller through finer classification.

Proposition 10. For arbitrary incomplete information system $S = (U, A)$, we have that

$$G(A) + E_{\mathrm{r}}(A) = \log_2|U|. \tag{12}$$

Proof. Let $S = (U, A)$ be an incomplete information system, $U/\mathrm{SIM}(A) = \{S_A(u_1), S_A(u_2), \ldots, S_A(u_{|U|})\}$. By the definitions 3 and 4

$$
\begin{aligned}
G(A) &= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_A(u_i)|}{|U|} \\
&= -\sum_{i=1}^{|U|} \frac{1}{|U|} (\log_2 |S_A(u_i)| - \log_2 |U|) \\
&= -\left(-\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_i)|} \right) + \log_2 |U| \sum_{i=1}^{|U|} \frac{1}{|U|} \\
&= -E_{\mathrm{r}}(A) + \log_2 |U|.
\end{aligned}
$$

i.e.

$$G(A) + E_{\mathrm{r}}(A) = \log_2 |U|.$$

This completes the proof. □

Example 3. Continued from example 2, by computing, it follows that

$$
\begin{aligned}
G(A) &= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_A(u_i)|}{|U|} \\
&= -\frac{1}{5} \left(\log_2 \frac{1}{5} + \log_2 \frac{1}{5} + \log_2 \frac{2}{5} + \log_2 \frac{2}{5} + \log_2 \frac{1}{5} \right) \\
&= \log_2 5 - \frac{2}{5}.
\end{aligned}
$$

and

$$
\begin{aligned}
E_{\mathrm{r}}(A) &= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_A(u_i)|} \\
&= -\frac{1}{5} \left(\log_2 \frac{1}{1} + \log_2 \frac{1}{1} + \log_2 \frac{1}{2} + \log_2 \frac{1}{2} + \log_2 \frac{1}{1} \right) \\
&= \frac{2}{5}.
\end{aligned}
$$

It is clear that $G(A) + E_{\mathrm{r}}(A) = \log_2 5 = \log_2 |U|$, i.e. the sum of the granulation and the rough entropy of knowledge A is the constant $\log_2 |U|$.

5. Two inequalities about the measures *GK*, *G*, *E* and E_{r}

By using Jensen's inequality two important inequalities about knowledge granulation GK, granularity measure G, rough entropy E_{r} and information entropy E are established in this section.

PROPOSITION 11. For arbitrary incomplete information system $S = (U, A)$, we have the following inequalities

$$-\log_2 GK(A) \leq G(A) \tag{13}$$

$$E_r(A) \leq \log_2(|U|(1 - E(A))). \tag{14}$$

Proof. Since $f(x) = -\log_2(x)$ is twice differentiable with

$$f''(x) = \frac{k}{x^2}$$

$f(x)$ is a convex function as $k = \log_2 e > 0$. By Definitions 2 and 3 and Jensen's inequality we have that

$$\begin{aligned}
-\log_2 GK(A) &= -\log_2\left(\frac{1}{|U|^2}\sum_{i=1}^{|U|}|S_A(u_i)|\right) \\
&= -\log_2\left(\sum_{i=1}^{|U|}\frac{1}{|U|}\frac{|S_A(u_i)|}{|U|}\right) \\
&\leq \sum_{i=1}^{|U|}\frac{1}{|U|}\left(-\log_2\frac{|S_A(u_i)|}{|U|}\right) \\
&= -\sum_{i=1}^{|U|}\frac{1}{|U|}\left(\log_2\frac{|S_A(u_i)|}{|U|}\right) \\
&= G(A).
\end{aligned}$$

Therefore, equation (13) holds.

By using the equalities $GK(A) + E(A) = 1$ and $G(A) + E_r(A) = \log_2|U|$, the formula (14) can be easily obtained. This completes the proof. □

Remark. It should be noted that the formulae (13) and (14) also hold for arbitrary complete information system.

Example 4. Continued from examples 2 and 3. We have that $E(A) = 8/25$, $GK(A) = 7/25$,

$$G(A) = \log_2 5 - \frac{2}{5}$$

and

$$E_r(A) = \frac{2}{5}.$$

From $(7/5)^5 \geq 4$, it is easy to conclude that

$$-\log_2\frac{7}{25} \leq \log_2 5 - \frac{2}{5}$$

i.e. $-\log_2 GK(A) \leq G(A)$.

From

$$E_r(A) = \frac{2}{5} \le \log_2 \frac{17}{5} = \log_2 5\left(1 - \frac{8}{25}\right) = \log_2(|U|(1 - E(A)))$$

it follows that $E_r(A) \le \log_2(|U|(1\text{-}E(A)))$.

6. Conclusions

In this paper, the concepts of information entropy, rough entropy and knowledge granulation in incomplete information systems are introduced, their important properties are given, the relationships among those concepts are established. The relationship between the information entropy $E(A)$ and the knowledge granulation $GK(A)$ of knowledge A can be expressed as $E(A) + GK(A) = 1$, the relationship between the granularity measure $G(A)$ and the rough entropy $E_r(A)$ of knowledge A can be expressed as $G(A) + E_r(A) = \log_2|U|$. Furthermore, two inequalities $-\log_2 GK(A) \le G(A)$ and $E_r(A) \le \log_2(|U|(1 - E(A)))$ about the measures GK, G, E and E_r are obtained. Information entropy, rough entropy and knowledge granulation characterize the significance of knowledge in different ways, and make more profound explanation. These results have a wide variety of applications, such as measuring the significance of attributes, constructing decision trees and measuring uncertainties of rules, etc. They will play a significant role in further researches in incomplete information systems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos 70471003, 60275019, 60573074), the national 863 plan project (No 2004AA115460), the foundation of doctoral program research of the Education Ministry of China (No 20050108004), the Natural Science Foundation of Shanxi, China (No 20041040) and the top scholar foundation of Shanxi, China.

References

T. Beaubouef, F.E. Petry and G. Arora, "Information-theoretic measures of un-certainty for rough sets and rough relational databases", *Inf. Sci.*, 109, pp. 535–563, 1998.

F.E. Chakik, A. Shahine, J. Jaam and A. Hasnah, "An approach for constructing complex discriminating surfaces based on bayesian interference of the maximum entropy", *Inf. Sci.*, 163, pp. 275–291, 2004.

I. Düntsch and G. Gediga, "Uncertainty measures of rough set prediction", *Artificial Intelligence*, 106, pp. 109–137, 1998.

J. Klir and M.J. Wierman, *Uncertainty Based Information: Elements of Generalized Information Theory*, New York: Physica-Verlag, 1999.

M. Kryszkiewicz, "Rough set approach to incomplete information systems", *Inf. Sci.*, 112, pp. 39–49, 1998.

M. Kryszkiewicz, "Rule in incomplete information systems", *Inf. Sci.*, 113, pp. 271–292, 1999.

Y. Leung and D.Y. Li, "Maximal consistent block technique for rule acquisition in incomplete information systems", *Inf. Sci.*, 153, pp. 85–106, 2003.

J.Y. Liang, K.S. Chin, C.Y. Dang and C.M.Y. Richard, "A new method for measuring uncertainty and fuzziness in rough set theory", *Int. J. Gen. Systems*, 31(4), pp. 331–342, 2002.

J.Y. Liang and Z.Z. Shi, "The information entropy, rough entropy and knowledge granulation in rough set theory", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(1), pp. 37–46, 2004.

J.Y. Liang and Z.B. Xu, "The algorithm on knowledge reduction in incomplete information systems", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(1), pp. 95–103, 2002.

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

J.S. Mi, W.Z. Wu and W.X. Zhang, “Approaches to knowledge reduction based on variable precision rough set model”, *Inf. Sci.*, 159, pp. 255–272, 2004.
Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Dordrecht: Kluwer Academic Publishers, 1991.
Z. Pawlak, J.W. Grzymala-Busse, R. Slowiski and W. Ziarko, “Rough sets”, *Comm. ACM*, 38(11), pp. 89–95, 1995.
C.E. Shannon, “The mathematical theory of communication”, *The Bell System Technical Journal*, 27(3 and 4), pp. 373–423, 1948.
M.J. Wierman, “Measuring uncertainty in rough set theory”, *Int. J. Gen. Systems*, 28(4), pp. 283–297, 1999.
W.X. Zhang, W.Z. Wu, J.Y. Liang and D.Y. Li, *Theory and Method of Rough Sets*, Beijing: Science Press, 2001.

Jiye Liang is a professor in the school of Computer & Information Technology, Shanxi University, in China. He received his PhD in Information Science in Xi’an Jiaotong University, and was engaged on Postdoctoral research about computer science and technology in the Institute of Computing Technology, the Chinese Academy of Sciences. Dr Liang has published over 60 technical papers in the field of rough set theory, information system and granular computing. His current research interests are intelligence information system, soft computing and granular computing.

Zhongzhi Shi is Professor at the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, IEEE Senior member, AAAI and ACM member. He is Chair of WG 12.2 of IFIP. He also serves as Vice President of the Chinese Association for Artificial Intelligence. He received the 2nd Grade National Award of Science and Technology Progress in 2002.In 1998 and 2001 he received the 2nd Grade Award of Science and Technology Progress from the Chinese Academy of Sciences. His research interests include intelligence science, multiagent systems, Semantic Web. He published 11 books, edited 11 books and more than 300 technical papers. His most recent books are *Intelligence Science* and *Knowledge Discovery* in Chinese. Homepages are available online at http://www.intsci.ac.cn/shizz.

Deyu Li received the MS degree in mathematics from Shanxi University in 1998 and PhD degree in computing mathematics in 2002 from Xi’an Jiaotong University, China. From November 2001 to May 2002, he worked as a research fellow at Department of Geography and Resource Management, Chinese University of Hong Kong. From January 2003 to January 2005, he was a postdoctoral researcher in State Key Laboratory of Intelligent Technology and System, Tsinghua University, China. He is now a professor at School of Computer and Information Technology, Shanxi University. He has published more than 20 academic papers on computing intelligence and mathematics. His current research interests include rough set theory and granular computing.