International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 12, No. 1 (2004) 37–46

World Scientific
www.worldscientific.com

THE INFORMATION ENTROPY, ROUGH ENTROPY AND KNOWLEDGE GRANULATION IN ROUGH SET THEORY

JIYE LIANG*,†,‡ and ZHONGZHI SHI*,§

*Key Laboratory of Intelligent Information Processing
Institute of Computing Technology
The Chinese Academy of Sciences
Beijing, 100080, People's Republic of China
†Department of Computer Science, Shanxi University
Taiyuan, 030006, People's Republic of China
‡ljy@sxu.edu.cn
§shizz@ics.ict.ac.cn

Received March 2003
Revised September 2003

Rough set theory is a relatively new mathematical tool for use in computer applications in circumstances which are characterized by vagueness and uncertainty. In this paper, we introduce the concepts of information entropy, rough entropy and knowledge granulation in rough set theory, and establish the relationships among those concepts. These results will be very helpful for understanding the essence of concept approximation and establishing granular computing in rough set theory.

Keywords: Rough sets; information entropy; rough entropy; knowledge granulation.

1. Introduction

Rough set theory, introduced by Z.Pawlak[1,2], is a relatively new soft computing tool for the analysis of a vague description of objects. The adjective vague, referring to the quality of information, means inconsistency or ambiguity which follows from information granulation. The rough sets 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, called elementary sets, that can be used to build knowledge about a real or abstract world. The use of the indiscernibility relation results in information granulation.

The entropy of a system as defined by Shannon [3] 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 ([4-5]) 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 in [6]. Unlike the logarithmic behavior of Shannon entropy, the gain function considered there possesses the complement nature. Especially, Wierman in [7] presented a well justified measure of uncertainty, 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.

In this paper, we introduce the concepts of information entropy, rough entropy and knowledge granulation in rough set theory, and establish the relationships among those concepts. These results will be very helpful for understanding the essence of concept approximation and establishing granular computing in rough set theory.

2. Information entropy and knowledge granulation

Let $K = (U, R)$ be an approximation space, where U: a non-empty, finite set called the universe; R: a partition of U, or an equivalence relation (i.e., indiscernibility relation) on U.

An approximation space $K = (U, R)$ can be regarded as a knowledge base about U.

Let

$$R = \{R_1, R_2, \cdots, R_m\}. \tag{1}$$

Of particular interest is the discrete partition,

$$\hat{R}(U) = \{\{x\}|x \in U\}, \tag{2}$$

and the indiscrete partition,

$$\check{R}(U) = \{U\}, \tag{3}$$

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

Given a partition R, and a subset $X \subseteq U$, we can define a lower approximation of X in U and a upper approximation of X in U by the following:

$$\underline{R}X = \bigcup\{R_i \in R|R_i \subseteq X\}, \tag{4}$$

and

$$\overline{R}X = \bigcup\{R_i \in R|R_i \cap X \neq \emptyset\}. \tag{5}$$

Both lower and upper approximation are unions of some equivalence classes. More precisely, the lower approximation $\underline{R}X$ is the union of those equivalence classes which are subsets of X. The upper approximation $\overline{R}X$ is the union of those equivalence classes which have a non-empty intersection with X.

The R-positive region of X is $POS_R(X) = \underline{R}X$, the R-negative region of X is $NEG_R(X) = U - \overline{R}X$, and the boundary or R-borderline region of X is $BN_R(X) = \overline{R}X - \underline{R}X$. X is called R-definable if and only if $\underline{R}X = \overline{R}X$. Otherwise, $\underline{R}X \neq \overline{R}X$ and X is rough with respect to R.

Definition 2.1. ([7]) Let $K = (U, R)$ be an approximation space, and R a partition of U. A measure of uncertainty in rough set theory is defined by

$$G(R) = -\sum_{i=1}^{m} \frac{|R_i|}{|U|} \log_2 \frac{|R_i|}{|U|}, \tag{6}$$

where $G : R \to [0, \infty)$ is a function from R, the set of all partitions of non-empty finite sets, to the non-negative real number, and $|U|$ is the cardinality of U. This granularity measure, G, measures the uncertainty associated with the prediction of outcomes where elements of each partition set R_i are indistinguishable.

If $R = \hat{R}$, then information measure of knowledge R achieves maximum value $\log_2 |U|$.

If $R = \check{R}$, then information measure of knowledge R achieves minimum value 0.

Obviously, when R is a partition of U, or an equivalence relation on U, we have that $0 \leq G(R) \leq \log_2 |U|$.

Now we define a partial order on all partition sets of U. Let P and Q be partitions of a finite set U, and we define the partition Q is coarser than the partition P (or the partition P is finer than the partition Q), $P \preceq Q$, between partitions by

$$P \preceq Q \Leftrightarrow \forall P_i \in P, \exists Q_j \in Q \to P_i \subseteq Q_j. \tag{7}$$

If $P \preceq Q$ and $P \neq Q$ then we say that Q is strictly coarser than P (or P is strictly finer than Q) and write $P \prec Q$.

Proposition 1.1. ([7]) Let P and Q be two partitions of finite set U. If $P \prec Q$, then $G(Q) < G(P)$.

Proposition 1.1 states that information measure of knowledge increases as the classes become smaller through finer partitioning.

If $p = (p_1, p_2, \cdots, p_n)$ is a finite probability distribution, then its Shannon entropy ([3]) is given by

$$S(p) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{8}$$

Let

$$p_i = \frac{|R_i|}{\sum_{j=1}^{m} |R_j|} = \frac{|R_i|}{|U|},$$

and it turns out that $p = (p_1, p_2, \cdots, p_m)$ is a probability distribution on R. Hence

$$G(R) = S(p). \tag{9}$$

The Hartley measure ([8]) of uncertainty for finite set X is

$$H(X) = \log_2 |X|. \tag{10}$$

The relationship between the granularity measure and the Hartley measure is as follows ([7]):

$$G(R) = H(U) - \sum_{i=1}^{m} \frac{|R_i|}{|U|} H(R_i). \tag{11}$$

We introduce a new definition of information entropy in rough set theory as follows.

Definition 2.2.([6]) Let $K = (U, R)$ be an approximation space, and R be a partition of U. An information entropy of knowledge R for rough set theory is defined by

$$E(R) = \sum_{i=1}^{m} \frac{|R_i|}{|U|} \frac{|R_i^c|}{|U|} = \sum_{i=1}^{m} \frac{|R_i|}{|U|}(1 - \frac{|R_i|}{|U|}), \tag{12}$$

where R_i^c is the complement of R_i, i.e., $R_i^c = U - R_i$; $\frac{|R_i|}{|U|}$ represents the probability of equivalence class R_i within the universe U; $\frac{|R_i^c|}{|U|}$ denotes the probability of the complement of R_i within the universe U.

If $R = \hat{R}$, then information entropy of knowledge R achieves maximum value $1 - 1/|U|$.

If $R = \check{R}$, then information entropy of knowledge R achieves minimum value 0.

Obviously, when R is a partition of U, or an equivalence relation on U, we have that $0 \leq E(R) \leq 1 - 1/|U|$.

Proposition 2.1. ([6]) Let P and Q be two partitions of finite set U. If $P \prec Q$, then $E(Q) < E(P)$.

Proposition 2.1 states that information entropy of knowledge increases as the classes become smaller through finer partitioning.

Definition 2.3. Let $K = (U, R)$ be an approximation space, and R a partition of U. Granulation of knowledge R is defined by

$$GK(R) = \frac{1}{|U|^2} \sum_{i=1}^{m} |R_i|^2, \tag{13}$$

where $\sum_{i=1}^{m} |R_i|^2$ is the cardinality of the equivalence relation $\bigcup_{i=1}^{m} (R_i \times R_i)$ determined by R.

If $R = \hat{R}$, then granulation of knowledge R achieves minimum value $|U|/|U|^2 = 1/|U|$.

If $R = \check{R}$, then granulation of knowledge R achieves maximum value $|U|^2/|U|^2 = 1$.

Obviously, when R is a partition of U, or an equivalence relation on U, we have that $1/|U| \leq GK(R) \leq 1$. Knowledge granulation can represent discernibility ability of knowledge, the smaller $GK(R)$ is , the stronger its discernibility ability.

Proposition 2.2. Let P and Q be two partitions of finite set U. If $P \prec Q$, then $GK(P) < GK(Q)$.

Proof. Let $P = \{P_1, P_2, \cdots, P_m\}$, and $Q = \{Q_1, Q_2, \cdots, Q_n\}$. Since $P \prec Q$, we have that $m > n$ and there exists a partition $C = \{C_1, C_2, \cdots, C_n\}$ of $\{1, 2, \cdots, m\}$ such that

$$Q_j = \bigcup_{i \in C_j} P_i, \ j = 1, 2, \cdots, n.$$

Hence

$$\begin{aligned} GK(Q) &= \tfrac{1}{|U|^2} \textstyle\sum_{j=1}^{n} |Q_j|^2 \\ &= \tfrac{1}{|U|^2} \textstyle\sum_{j=1}^{n} |\bigcup_{i \in C_j} P_i|^2 \\ &= \tfrac{1}{|U|^2} \textstyle\sum_{j=1}^{n} (\sum_{i \in C_j} |P_i|)^2. \end{aligned}$$

From $m > n$ it follows that there exists $C_{j_0} \in C$ such that $|C_{j_0}| > 1$. Therefore

$$(\sum_{i \in C_{j_0}} |P_i|)^2 > \sum_{i \in C_{j_0}} |P_i|^2$$

and

$$(\sum_{i \in C_j, j \neq j_0} |P_i|)^2 \geq \sum_{i \in C_j, j \neq j_0} |P_i|^2.$$

Thus

$$GK(Q) > \frac{1}{|U|^2}\sum_{i=1}^{m}|P_i|^2 = GK(P).$$

i.e.,

$$GK(P) < GK(Q).$$

This completes the proof. □

Proposition 2.2 states that knowledge granulation decreases as the classes become smaller through finer partitioning.

Proposition 2.3. Let $K = (U, R)$ be an approximation space, and R be a partition of U, then the relationship between the information entropy and knowledge granulation is as follows:

$$E(R) = 1 - GK(R). \tag{14}$$

Proof. Let $R = \{R_1, R_2, \cdots, R_m\}$, then

$$\begin{aligned} E(R) &= \sum_{i=1}^{m} \tfrac{|R_i|}{|U|}(1 - \tfrac{|R_i|}{|U|}) \\ &= \sum_{i=1}^{m} \tfrac{|R_i|}{|U|} - \sum_{i=1}^{m} \tfrac{|R_i^2|}{|U|^2} \\ &= 1 - GK(R). \end{aligned}$$

This completes the proof. □

From proposition 2.2 we can obtain $E(R) + GK(R) = 1$, where $0 \leq E(R) \leq 1 - 1/|U|$ and $1/|U| \leq GK(R) \leq 1$.

Example 2.1. Let U ={medium, small, little, tiny, big, large, huge, enormous }. The equivalence relation R, i.e., a partition of U is defined as follows: R ={{medium}, {small, little, tiny}, {big, large}, {huge, enormous}}. By computing, it follows that

$$\begin{aligned} E(R) &= \sum_{i=1}^{m} \tfrac{|R_i|}{|U|}(1 - \tfrac{|R_i|}{|U|}) \\ &= \tfrac{1}{8}(1 - \tfrac{1}{8}) + \tfrac{3}{8}(1 - \tfrac{3}{8}) + \tfrac{2}{8}(1 - \tfrac{2}{8}) + \tfrac{2}{8}(1 - \tfrac{2}{8}) \\ &= \tfrac{23}{32} = 0.71875 \end{aligned}$$

and

$$\begin{aligned} GK(R) &= \tfrac{1}{|U|^2} \sum_{i=1}^{m} |R_i|^2 \\ &= \tfrac{1}{8^2}(1^2 + 3^2 + 2^2 + 2^2) \\ &= \tfrac{9}{32} = 0.28125. \end{aligned}$$

It is clear that $E(R) + GK(R) = 1$.

3. Granularity measure and rough entropy

The concept of rough entropy has been introduced in rough sets, rough relational databases and incomplete information systems [4,9]. Now we introduce a definition of rough entropy of knowledge in approximate space.

Definition 3.1. Let $K = (U, R)$ be an approximate space, and R a partition of U. The rough entropy $E_r(R)$ of knowledge R is defined by

$$E_r(R) = -\sum_{i=1}^{m} \frac{|R_i|}{|U|} \log_2 \frac{1}{|R_i|}, \tag{15}$$

where $|R_i|/|U|$ represents the probability of equivalence class R_i within the universe U, $1/|R_i|$ denotes the probability of one of the values in equivalence class R_i.

If $R = \hat{R}$, then the rough entropy of knowledge R achieves minimum value 0.

If $R = \check{R}$, then the rough entropy of knowledge R achieves maximum value $\log_2 |U|$.

Obviously, when R is a partition of U, or an equivalence relation on U, we have that $0 \leq E_r(R) \leq log_2|U|$.

Proposition 3.1. Let P and Q be two partitions of finite set U. If $P \prec Q$, then $E_r(P) < E_r(Q)$.

Proof. Let $P = \{P_1, P_2, \cdots, P_m\}$, and $Q = \{Q_1, Q_2, \cdots, Q_n\}$. Since $P \prec Q$, we have that $m > n$ and there exists a partition $C = \{C_1, C_2, \cdots, C_n\}$ of $\{1, 2, \cdots, m\}$ such that

$$Q_j = \bigcup_{i \in C_j} P_i, \ j = 1, 2, \cdots, n.$$

Hence

$$\begin{aligned} E_r(Q) &= -\sum_{j=1}^{n} \tfrac{|Q_j|}{|U|} \log_2 \tfrac{1}{|Q_j|} \\ &= \tfrac{1}{|U|} \sum_{j=1}^{n} |Q_j| \log_2 |Q_j| \\ &= \tfrac{1}{|U|} \sum_{j=1}^{n} |\bigcup_{i \in C_j} P_i| \log_2 |\bigcup_{i \in C_j} P_i| \\ &= \tfrac{1}{|U|} \sum_{j=1}^{n} (\sum_{i \in C_j} |P_i| \log_2 (\sum_{i \in C_j} |P_i|)) \end{aligned}$$

From $m > n$ it follows that there exists $C_{j_0} \in C$ such that $|C_{j_0}| > 1$. Therefore

$$\sum_{i \in C_{j_0}} |P_i| \log_2 (\sum_{i \in C_{j_0}} |P_i|) > \sum_{i \in C_{j_0}} |P_i| \log_2 |P_i|$$

and

$$\sum_{i\in C_j, j\neq j_0} |P_i| \log_2 (\sum_{i\in C_j, j\neq j_0} |P_i|) \geq \sum_{i\in C_j, j\neq j_0} |P_i| \log_2 |P_i|.$$

Thus

$$\begin{aligned} E_r(Q) &> \frac{1}{|U|}\sum_{i=1}^{m} |P_i| \log_2 |P_i| \\ &= -\sum_{i=1}^{m} \frac{|P_i|}{|U|} \log_2 \frac{1}{|P_i|} \\ &= E_r(P). \end{aligned}$$

i.e.,

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

This completes the proof. □

Proposition 3.1 states that the rough entropy of knowledge decreases as the classes become smaller through finer partitioning.

Proposition 3.2. Let $K = (U, R)$ be an approximate space, and R a partition of U. Then the relationship between the granularity measure and the rough entropy of knowledge is as follows :

$$G(R) + E_r(R) = log_2|U|. \tag{16}$$

Proof. Let $R = \{R_1, R_2, \cdots, R_m\}$, then

$$\begin{aligned} G(R) &= -\sum_{i=1}^{m} \frac{|R_i|}{|U|} \log_2 \frac{|R_i|}{|U|} \\ &= -\sum_{i=1}^{m} \frac{|R_i|}{|U|}(\log_2 |R_i| - log_2|U|) \\ &= -(-\sum_{i=1}^{m} \frac{|R_i|}{|U|} \log_2 \frac{1}{|R_i|}) + \log_2 |U| \sum_{i=1}^{m} \frac{|R_i|}{|U|} \\ &= -E_r(R) + log_2|U|. \end{aligned}$$

i.e.,

$$G(R) + E_r(R) = log_2|U|.$$

This completes the proof. □

Example 3.1. Continued from Example 2.1, by computing, it follows that

$$\begin{aligned} G(R) &= -\sum_{i=1}^{m} \frac{|R_i|}{|U|} \log_2 \frac{|R_i|}{|U|} \\ &= -[\tfrac{1}{8}log_2\tfrac{1}{8} + \tfrac{3}{8}log_2\tfrac{3}{8} + \tfrac{2}{8}log_2\tfrac{2}{8} + \tfrac{2}{8}log_2\tfrac{2}{8}] \\ &= \tfrac{5}{2} - \tfrac{3}{8}log_2 3 \end{aligned}$$

and

$$
\begin{aligned}
E_r(R) &= -\sum_{i=1}^{m} \frac{|R_i|}{|U|} \log_2 \frac{1}{|R_i|} \\
&= -[\tfrac{1}{8}log_2\tfrac{1}{1} + \tfrac{3}{8}log_2\tfrac{1}{3} + \tfrac{2}{8}log_2\tfrac{1}{2} + \tfrac{2}{8}log_2\tfrac{1}{2}] \\
&= \tfrac{1}{2} + \tfrac{3}{8}log_2 3.
\end{aligned}
$$

It is clear that $G(R) + E_r(R) = 3 = log_2|U|$.

4. Conclusions

In this paper, the concepts of information entropy, rough entropy and knowledge granulation in rough set theory have been introduced, their important properties are given, the relationships among those concepts are established, and we have shown that the relationship between the information entropy $E(R)$ and the knowledge granulation $GK(R)$ of knowledge R is $E(R) = 1 - GK(R)$, the relationship between the granularity measure $G(R)$ and the rough entropy $E_r(R)$ of knowledge R is $G(R) + E_r(R) = log_2|U|$. These results will play a significance role in further research on rough set theory.

Acknowledgements

This work was supported by the national natural science foundation of China (No.60275019) and the natural science foundation of Shanxi, China (No.20031036).

References

1. Z. Pawlak, *Rough sets: Theoretical Aspects of Reasoning about Data* (Kluwer Academic Publishers, Dordrecht, 1991).
2. Z. Pawlak, J.W. Grzymala-Busse, R. Słowiński and W. Ziarko, "Rough sets", *Comm. ACM* **38**(1995) 89-95.
3. C.E. Shannon, "The mathematical theory of communication", *The Bell System Technical Journal* **27**(3 and 4)(1948) 373-423.
4. T.Beaubouef, F.E.Petry and G.Arora,"Information-theoretic measures of uncertainty for rough sets and rough relational databases", *Information Sciences* **109**(1998) 535-563.
5. I. Düntsch and G. Gediga,"Uncertainty measures of rough set prediction ", *Artificial Intelligence* **106**(1998) 109-137.
6. J.Y. Liang, K.S Chin, C.Y. Dang and C.M.YAM. Richard, "A new method for measuring uncertainty and fuzziness in rough set theory", *International Journal of General Systems* **31** (2002) 331-342.
7. M.J. Wierman, "Measuring uncertainty in rough set theory", *International Journal of General Systems* **28** (1999) 283-297.
8. R.V.L. Hartley, "Transmission of information", *The Bell Systems Technical Journal* **7**(1928) 535-563.

9. 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** (2002) 95-103.