International Journal of General Systems, 2002 VOL. 31 (4), pp. 331–342

Taylor & Francis
Taylor & Francis Group

A NEW METHOD FOR MEASURING UNCERTAINTY AND FUZZINESS IN ROUGH SET THEORY

JIYE LIANG[a,*], K.S. CHIN[b], CHUANGYIN DANG[b] and RICHARD C.M. YAM[b]

[a]*Department of Computer Science, Shanxi University, Taiyuan 030006, People's Republic of China;*
[b]*Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong, People's Republic of China*

(Received 21 February 2001; In final form 15 February 2002)

Based on the complement behavior of information gain, a new definition of information entropy is proposed along with its justification in rough set theory. Some properties of this definition imply those of Shannon's entropy. Based on the new information entropy, conditional entropy and mutual information are then introduced and applied to knowledge bases. The new information entropy is proved to also be a fuzzy entropy.

Keywords: Rough sets; Information entropy; Uncertainty; Fuzziness; Rough classification; Data analysis

INTRODUCTION

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 (Düntsch and Gediga, 1998; Beaubouef *et al.*, 1998; Wierman, 1999; Liang and Xu, 2000; Liang *et al.*, 2000) have used Shannon's concept and its variants to measure uncertainty in rough set theory. But Shannon's entropy is not a fuzzy entropy, and cannot measure the fuzziness in rough set theory.

This paper introduces a new definition for information entropy in rough set theory. Unlike the logarithmic behavior of Shannon's entropy, the gain function considered here possesses the complement nature. Some important properties of this definition are also derived. Based on the new concept, conditional entropy and mutual information are then introduced and applied to knowledge bases. The new measure of information is also proved to be a fuzzy entropy, and can be used to measure the fuzziness of rough set and rough classification.

MEASURE OF UNCERTAINTY IN ROUGH SET THEORY

Rough set theory (Pawlak, 1991) has become well established as a mechanism for uncertainty management in a wide variety of applications related to Artificial

*Corresponding author.

ISSN 0308-1079 print/ISSN 1563-5104 online © 2002 Taylor & Francis Ltd
DOI: 10.1080/0308107021000013635

Intelligence (Dubois and Prade, 1990; Slowiński, 1992; Pawlak *et al.*, 1995; Lingras and Yao, 1998).

Let

$$K = (U, R)$$

be an approximation space, where U is a non-empty, finite set called the universe; R is a partition of U, or an equivalence relation on U.

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

Let

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

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:

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

and

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

The R-positive region of X is $\mathrm{POS}_R(X) = \underline{R}X$, the R-negative region of X is $\mathrm{NEG}_R(X) = U - \overline{R}X$, and the boundary or R-borderline region of X is $\mathrm{BN}_R(X) = \overline{R}X - \underline{R}X$. 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 1 (Wierman, 1999) 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 $G : R \rightarrow [0, \infty)$ 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 $p = (p_1, p_2, \ldots, p_n)$ is a finite probability distribution, then its Shannon entropy (Shannon, 1948; Klir and Wierman, 1998) is given by

$$S(p) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{7}$$

Let

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

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

$$G(R) = S(p). \tag{8}$$

The Hartley measure (Hartley, 1928) of uncertainty for finite set X is

$$H(X) = \log_2|X|. \tag{9}$$

The relationship between the granularity measure and the Hartley measure is as follows (Wierman, 1999):

$$G(R) = H(U) - \sum_{i=1}^{m} \frac{|R_i|}{|U|} H(R_i). \tag{10}$$

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

DEFINITION 2 Let $K = (U, R)$ be an approximation space, and R be a partition of U. Information entropy for rough set theory is defined by

$$E(R) = \sum_{i=1}^{m} \frac{|R_i|}{|U|} \frac{|R_i^c|}{|U|} = \sum_{i=1}^{m} \frac{|R_i|}{|U|} \left(1 - \frac{|R_i|}{|U|}\right) \tag{11}$$

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

$$P \preceq Q \Leftrightarrow \forall P_i \in P, \exists Q_j \in Q \rightarrow P_i \subseteq Q_j. \tag{12}$$

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 (CARDINALITY) If P and Q are partitions of U with $|P| = |Q|$ and there exists a one-to-one, onto function $h : P \rightarrow Q$ such that

$$|h(P_i)| = |P_i|,$$

then

$$E(P) = E(Q).$$

Proposition 1 states that the uncertainty is invariant with respect to different partitions of U that are size-isomorphic.

We first prove the following lemma in order to derive other propositions later.

LEMMA 1 Let p be a finite probability distribution in U. Let $E(p) = \sum_{x \in U} p(x) \times (1 - p(x)) = 1 - \sum_{x \in U} p^2(x)$. Then

(1) $0 \leq E(p) \leq 1 - 1/n(|U| = n)$;
(2) $E(p) = 1 - 1/n$ iff $p(x) = 1/n(x \in U)$.

Proof Let

$$H(\lambda) = \sum_{x \in U} p^2(x) + \lambda \left(\sum_{x \in U} p(x) - 1 \right).$$

Since

$$\begin{cases} H'_{\lambda}(\lambda) = \sum_{x \in U} p(x) - 1 = 0 \\ H'_{p(x)}(\lambda) = 2p(x) + \lambda = 0 \quad (x \in U) \end{cases}$$

we know that $p(x) = 1/n (x \in U)$. So the minimum value $1/n$ of $\sum_{x \in U} p^2(x)$ can be achieved only under the restriction $\sum_{x \in U} p(x) = 1$ when $p(x) = 1/n (x \in U)$. □

PROPOSITION 2 (MONOTONICITY) If X and Y are finite sets and $|X| < |Y|$, then

$$E(\hat{R}(X)) < E(\hat{R}(Y)).$$

Proof Let $p(x) = |\{x\}|/|X| = 1/|X| (x \in X)$, and $p(y) = |\{y\}|/|Y| = 1/|Y| (y \in Y)$. From Lemma 1, we have that $E(\hat{R}(X)) = 1 - 1/|X|$ and $E(\hat{R}(Y)) = 1 - 1/|Y|$. Since $|X| < |Y|$, it follows that $1 - 1/|X| < 1 - 1/|Y|$, i.e. $E(\hat{R}(X)) < E(\hat{R}(Y))$. □

PROPOSITION 3 (ROUGHNESS MONOTONICITY) Let P and Q be two partitions of finite set U. If $P \prec Q$, then $E(Q) < E(P)$.

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

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

Hence

$$E(Q) = \sum_{j=1}^{n} \left(\frac{|Q_i|}{|U|} \right) \left(1 - \frac{|Q_i|}{|U|} \right) = 1 - \frac{1}{|U|^2} \sum_{j=1}^{n} |Q_j|^2 = 1 - \frac{1}{|U|^2} \sum_{j=1}^{n} \left| \bigcup_{i \in C_j} P_i \right|^2$$

$$= 1 - \frac{1}{|U|^2} \sum_{j=1}^{n} \left(\sum_{i \in C_j} |P_i| \right)^2.$$

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

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

and

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

Thus

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

From Proposition 3, it is clear that the information entropy E increases monotonically as the granularity of information becomes smaller through finer partitions. □

COROLLARY 1 Let P and Q be two partitions of finite set U. If $P \preceq Q$ and $E(P) = E(Q)$, then $P = Q$.

From Lemma 1 and Proposition 3, one can obtain immediately the following propositions.

PROPOSITION 4 (MAXIMUM) The maximum of the information entropy E for any finite set U is $1 - (1/|U|)$. This value is achieved only by the discrete partition $\hat{R}(U)$.

PROPOSITION 5 (MINIMUM) The minimum of the information entropy E for any finite set U is 0. This value is achieved only by the indiscrete partition $\check{R}(U)$.

INFORMATION MEASURE BETWEEN KNOWLEDGE BASES

Let U be the universal set, and $K_1 = (U, P)$ and $K_2 = (U, Q)$ be two knowledge bases about U, where $P = \{P_1, P_2, \ldots, P_m\}$, and $Q = \{Q_1, Q_2, \ldots, Q_n\}$.

DEFINITION 3 Conditional entropy $E(Q|P)$ of Q about P is defined by

$$E(Q|P) = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|Q_i \cap P_j|}{|U|}\frac{|Q_i^c - P_j^c|}{|U|}. \tag{13}$$

DEFINITION 4 Mutual information $E(Q; P)$ of Q and P is defined by

$$E(Q; P) = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|Q_i \cap P_j|}{|U|}\frac{|Q_i^c \cap P_j^c|}{|U|}. \tag{14}$$

PROPOSITION 6 Let U be the universal set, and $K_1 = (U, P)$ and $K_2 = (U, Q)$ be two knowledge bases about U, then

$$E(Q; P) = E(Q) - E(Q|P). \tag{15}$$

Proof Since $Q_i^c = (Q_i^c \cap P_j^c) \cup (Q_i^c - P_j^c)$, we have that

$$E(Q) = \sum_{i=1}^{n}\frac{|Q_i|}{|U|}\frac{|Q_i^c|}{|U|} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|Q_i \cap P_j|}{|U|}\frac{|Q_i^c|}{|U|}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|Q_i \cap P_j|}{|U|}\frac{|(Q_i^c \cap P_j^c) \cup (Q_i^c - P_j^c)|}{|U|} = E(Q; P) + E(Q|P).$$

Hence

$$E(Q; P) = E(Q) - E(Q|P).$$ □

PROPOSITION 7 Let U be the universal set, $K_1 = (U, P)$ and $K_2 = (U, Q)$ be two knowledge bases about U, and D be a decision on U (i.e. a partition of U). If $P \prec Q$, then $E(D; P) \geq E(D; Q)$.

Proof Let

$$P = \{P_1, P_2, \ldots, P_m\}, \quad Q = \{Q_1, Q_2, \ldots, Q_n\},$$

and

$$D = \{d_1, d_2, \ldots, d_r\}.$$

Since $P \prec Q$, we have that $m > n$ and there exists a partition $C = \{C_1, C_2, \ldots, C_n\}$ of $\{1, 2, \ldots, m\}$ such that

$$Q_j = \bigcup_{k \in C_j} P_k, \quad j = 1, 2, \ldots, n.$$

Hence

$$\begin{aligned}
E(D; Q) &= \sum_{i=1}^{r} \sum_{j=1}^{n} \frac{|d_i \cap Q_j|}{|U|} \frac{|d_i^c \cap Q_j^c|}{|U|} = \sum_{i=1}^{r} \sum_{j=1}^{n} \frac{|d_i \cap Q_j|}{|U|} \frac{|(U - (d_i \cup Q_j)|}{|U|} \\
&= \sum_{i=1}^{r} \sum_{j=1}^{n} \frac{\left| d_i \cap \bigcup_{k \in C_j} P_k \right|}{|U|} \frac{\left| U - \left(d_i \cup \bigcup_{k \in C_j} P_k \right) \right|}{|U|} \\
&= \sum_{i=1}^{r} \sum_{j=1}^{n} \frac{\left(\sum_{k \in C_j} |d_i \cap P_k| \right)}{|U|} \frac{\left| U - \left(d_i \cup \bigcup_{k \in C_j} P_k \right) \right|}{|U|} \\
&\leq \sum_{i=1}^{r} \sum_{k=1}^{m} \frac{|d_i \cap P_k|}{|U|} \frac{|U - (d_i \cup P_k)|}{|U|} = \sum_{i=1}^{r} \sum_{k=1}^{m} \frac{|d_i \cap P_k|}{|U|} \frac{|d_i^c \cap P_k^c|}{|U|} = E(D; P).
\end{aligned}$$

□

Example 1 Reverse relation of Proposition 7 cannot be established in general.

Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Assume that

$$P = \{\{1, 5\}, \{2, 3, 4, 6, 7\}, \{8, 9, 10\}\}$$
$$Q = \{\{1, 3, 4\}, \{2, 5, 6\}, \{7, 8, 9, 10\}\},$$

and

$$D = \{\{1, 3, 5, 8, 9\}, \{2, 4, 6, 7, 10\}\}.$$

It is easily computed that

$$E(D; P) = 0.38, \quad E(D; Q) = 0.34,$$

i.e.

$$E(D; P) > E(D; Q).$$

However, we have that $P \not\prec Q$.

PROPOSITION 8 Let U be the universal set, and $K_1 = (U, P)$ and $K_2 = (U, Q)$ be two knowledge bases about U. Then $P \preceq Q$ if and only if $E(Q|P) = 0$.

Proof

i) Suppose that $P \preceq Q$. From $P \preceq Q$, it follows that $Q_i \cap P_j = \emptyset$ or $P_j \subseteq Q_i$ for $\forall P_j \in P$ and $\forall Q_i \in Q$. Thus, $|Q_i \cap P_j|\,|P_j - Q_i| = 0$ for $\forall P_j \in P$ and $\forall Q_i \in Q$. Hence, we have that

$$E(Q|P) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|Q_i \cap P_j|}{|U|}\frac{|Q_i^c - P_j^c|}{|U|} = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|Q_i \cap P_j|}{|U|}\frac{|P_j - Q_i|}{|U|} = 0.$$

ii) Suppose that $E(Q|P) = 0$. We want to prove that $P \preceq Q$. Assume that $P \not\preceq Q$. Then there exists a $P_k \in P$ such that

$$P_k \not\subseteq Q_i, \quad \forall Q_i \in Q.$$

Let $\{Q_i \in Q | Q_i \cap P_k \neq \emptyset\} = \{Q_{i_1}, Q_{i_2}, \ldots, Q_{i_{k'}}\}$, where $k' > 1$. Then

$$|Q_{i_l} \cap P_k| > 0 \quad \text{and} \quad |P_k - Q_{i_l}| > 0, \quad l = 1, 2, \ldots, k'.$$

Hence

$$E(Q|P) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|Q_i \cap P_j|}{|U|}\frac{|Q_i^c - P_j^c|}{|U|} = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|Q_i \cap P_j|}{|U|}\frac{|P_j - Q_i|}{|U|}$$

$$\geq \sum_{i=1}^{n} \frac{|Q_i \cap P_k|}{|U|}\frac{|P_k - Q_i|}{|U|} \geq \sum_{l=1}^{k'} \frac{|Q_{i_l} \cap P_k|}{|U|}\frac{|P_k - Q_{i_l}|}{|U|} > 0.$$

This yields a contradiction. Therefore, $P \preceq Q$. □

MEASURE OF FUZZY INFORMATION OF ROUGH SET AND ROUGH CLASSIFICATION

Let U be the universal set, $F(U)$ the class of all fuzzy sets of U, $\mu_A(x)$ the membership function of $A \in F(U)$, $\varphi(U)$ the class of all crisp sets of U, $[1/2]_U$ the fuzzy set of U for which $\mu_{[1/2]_U}(x) = 1/2$, $\forall x \in U$, and F a sub-class of $F(U)$ with (1) $\varphi(U) \subseteq F$, (2) $[1/2]_U \in F$, (3) $A, B \in F \Rightarrow A \cup B \in F$, $A^c \in F$, where $A^c \in F(U)$ is the complement of $A \in F(U)$, i.e. $\mu_{A^c}(x) = 1 - \mu_A(x)$, $\forall x \in U$.

The entropy of a fuzzy set is a measure of fuzziness of the fuzzy set. De Luca and Termini (1972) introduced the axiomatic construction of entropy of fuzzy sets and referred to Shannon's probability entropy. Liu (1992) systematically gave the axiomatic definitions of entropy, distance measure and similarity measure of fuzzy sets and discuss some basic relations between these measures.

DEFINITION 5 (LIU, 1992) A real function $e : F \rightarrow [0, +\infty)$ is called an entropy on F if e has the following properties:

(1) $e(D) = 0$, $\forall D \in \varphi(U)$;
(2) $e([1/2]_U) = \max_{A \in F} e(A)$;

(3) $\forall A, B \in F$, if $\mu_B(x) \geq \mu_A(x)$ when $\mu_A(x) \geq 1/2$, or $\mu_B(x) \leq \mu_A(x)$ when $\mu_A(x) \leq 1/2$, then $e(A) \geq e(B)$;
(4) $e(A^c) = e(A)$, $\forall A \in F$.

Let $U = \{x_1, x_2, \ldots, x_n\}$, and define

$$E(A) = \sum_{i=1}^{n} \mu_A(x_i)\,(1 - \mu_A(x_i)), \quad \forall A \in F. \tag{16}$$

Then E is an entropy on F.

In fact, we have the following:

(1) For $\forall D \in \varphi(U)$, we have $\mu_D(x) = 1$ or 0 for $\forall x \in U$. Hence, $E(D) = 0$.
(2) From $0 \leq \mu_A(x) \leq 1$, it follows that $\max_{A \in F}(\mu_A(x)(1 - \mu_A(x))) = \mu_{A_0}(x) \times (1 - \mu_{A_0}(x)) = 1/4$, where $A_0 \in F$ and $\mu_{A_0}(x) = 1/2$ for $\forall x \in U$. Hence, $E([1/2]_U) = \max_{A \in F} E(A)$.
(3) Let $A, B \in F$ be two arbitrary fuzzy sets. Let $f(\mu(x)) = \mu(x)(1 - \mu(x))$, where $0 \leq \mu(x) \leq 1$, and $x \in U$.

It can be easily proved that $f(\mu(x))$ is strictly increasing on $\mu(x) \in [0, 1/2]$; $f(\mu(x))$ is strictly decreasing on $\mu(x) \in [1/2, 1]$.

Assume that $\mu_B(x) \geq \mu_A(x)$ when $\mu_A(x) \geq 1/2$ or $\mu_B(x) \leq \mu_A(x)$ when $\mu_A(x) \leq 1/2$. From the property of $f(\mu(x))$, it follows that $f(\mu_A(x)) \geq f(\mu_B(x))$. Thus, $\sum_{i=1}^{n} f(\mu_A(x_i)) \geq \sum_{i=1}^{n} f(\mu_B(x_i))$, i.e. $E(A) \geq E(B)$.
(4) Let $A \in F$. Since $\mu_{A^c}(x) = 1 - \mu_A(x)$, it follows that $\mu_A(x)(1 - \mu_A(x)) = \mu_{A^c}(x) \times (1 - \mu_{A^c}(x))$. Therefore, $E(A^c) = E(A)$.

Summarizing above, we obtain that E is an entropy on F.

We remark that

$$S(A) = -\sum_{i=1}^{n} \mu_A(x_i) \log_2 \mu_A(x_i), \quad \forall A \in F$$

is not a fuzzy entropy.

DEFINITION 6 (LIU, 1992) Let e be an entropy on F. If, for $\forall A \in F$,

$$e(A) = e(A \cap D) + e(A \cap D^c), \quad \forall D \in \varphi(U) \tag{17}$$

then we call e a σ-entropy on F.

PROPOSITION 9 The entropy E is a σ-entropy on $F(U)$.

Proof Let $U = \{x_1, x_2, \ldots, x_n\}$. For $\forall A \in F(U)$ and $\forall D \in \varphi(U)$,

$$\begin{aligned} E(A \cap D) + E(A \cap D^c) &= \sum_{i=1}^{n} \mu_{A \cap D}(x_i)(1 - \mu_{A \cap D}(x_i)) + \sum_{i=1}^{n} \mu_{A \cap D^c}(x_i)(1 - \mu_{A \cap D^c}(x_i)) \\ &= \sum_{i=1}^{n} (\mu_{A \cap D}(x_i) + \mu_{A \cap D^c}(x_i)) - \sum_{i=1}^{n} ((\mu_{A \cap D}(x_i))^2 + (\mu_{A \cap D^c}(x_i))^2) \\ &= \sum_{i=1}^{n} \mu_A(x_i) - \sum_{i=1}^{n} (\mu_A(x_i))^2 = \sum_{i=1}^{n} \mu_A(x_i)(1 - \mu_A(x_i)) = E(A). \end{aligned}$$

Thus, E is a σ-entropy on $F(U)$.

Let $K = (U, R)$ be an approximation space, where R is an equivalence relation on U or a partition of U. Let $[x]_R$ denote the equivalence class of the relation R containing the element x. Then, for any non-null subset X of U, in terms of equivalence classes, the lower and upper approximations of X in K can be expressed, respectively, by

$$\underline{R}X = \{x \in U | [x]_R \subseteq X\}$$

and

$$\overline{R}X = \{x \in U | [x]_R \cap X \neq \emptyset\}.$$

For an element $x \in U$, the degree of rough belongings (Pawlak, 1991) of x in X is given by

$$\mu_X(x) = \frac{|X \cap [x]_R|}{|[x]_R|} \tag{18}$$

where $\mu_X(x) (0 \leq \mu_X(x) \leq 1)$ represents a vague concept.

This immediately induces a fuzzy set F_X^R of U given by $F_X^R = \{(x, \mu_X(x)) | x \in U\}$. □

DEFINITION 7 The measure of fuzziness of rough set X in an approximation space $K = (U, R)$ is defined by

$$E(F_X^R) = \sum_{i=1}^{n} \mu_X(x_i)(1 - \mu_X(x_i)) \tag{19}$$

where $|U| = n$.

PROPOSITION 10 The fuzziness of an exact set in an approximation space is 0.

Proof Let X be an exact set in an approximation space (U, R). Then $\underline{R}X = X = \overline{R}X$, and $\forall x \in X$, $\mu_X(x) = (|X \cap [x]_R|)/(|[x]_R|) = (|[x]_R|)/(|[x]_R|) = 1$. For each $x \in U - X$, $[x]_R \cap X = \emptyset$. Hence, for each $x \in U - X$, $\mu_X(x) = 0$. Thus, for each $x \in U$, $\mu_X(x) \times (1 - \mu_X(x)) = 0$, i.e. $E(F_X^R) = 0$. □

PROPOSITION 11 A rough set and its complement have the same fuzziness.

Proof Let X be a rough set in an approximation space (U, R), and X^c its complement. For $\forall x \in U$, we have that

$$\mu_X(x) + \mu_{X^c}(x) = \frac{|X \cap [x]_R|}{|[x]_R|} + \frac{|X^c \cap [x]_R|}{|[x]_R|} = \frac{|[x]_R|}{|[x]_R|} = 1,$$

i.e. $\mu_{X^c}(x) = 1 - \mu_X(x)$. Thus, for $\forall x \in U$, $\mu_X(x)(1 - \mu_X(x)) = \mu_{X^c}(x)(1 - \mu_{X^c}(x))$, i.e. $E(F_X^R) = E(F_{X^c}^R)$.

Let $C = \{C_1, C_2, \ldots, C_r\}$ be a classification of U, i.e. C be a partition of U. C_i are called class of C. Then

$$\underline{R}C = \{\underline{R}C_1, \underline{R}C_2, \ldots, \underline{R}C_r\}$$

and

$$\overline{R}C = \{\overline{R}C_1, \overline{R}C_2, \ldots, \overline{R}C_r\}$$

are, respectively, called the lower and upper approximations of C in K.

For an element $x \in U$, the degree of rough classification of x in C is given by

$$\mu_C(x) = \frac{|C_j \cap [x]_R|}{|[x]_R|}, \quad x \in C_j \tag{20}$$

where $\mu_C(x)(0 \leq \mu_C(x) \leq 1)$ represents a vague concept.

This immediately induces a fuzzy set F_C^R of U given by $F_C^R = \{(x, \mu_C(x))|x \in U\}$. □

DEFINITION 8 The measure of fuzziness of rough classification $C = \{C_1, C_2, \ldots, C_r\}$ in an approximation space $K = (U, R)$ is defined by

$$E(F_C^R) = \sum_{i=1}^{n} \mu_C(x_i)(1 - \mu_C(x_i)). \tag{21}$$

PROPOSITION 12 The fuzziness of an exact classification in an approximation space is 0.

Proof Let $C = \{C_1, C_2, \ldots, C_r\}$ be an exact classification in an approximation space (U, R), i.e. $\underline{R}C_i = \overline{R}C_i$, $i = 1, 2, \ldots, r$. Then, for $\forall x \in U$, there exists uniquely $C_j \in C$ such that $x \in C_j$, and $\mu_C(x) = (|C_j \cap [x]_R|)/(|[x]_R|) = (|[x]_R|)/(|[x]_R|) = 1$. Therefore, $E(F_C^R) = 0$. □

CONCLUSIONS

A new definition of information entropy based on the complement behavior of information gain has been proposed along with its justification in rough set theory. Based on this concept, conditional entropy and mutual information have been introduced. In particular, the new information entropy can measure both uncertainty and fuzziness in rough set theory. Now we are studying information measures in generalized rough set model (Slowiński and Vanderpooten, 2000) for data mining applications, which will be reported in another paper.

Acknowledgements

This work was supported by UGC of Hong Kong (No. 9350007), the national 863 plan of China (No. 2001AA115460, No. 2001AA114031), the national natural science foundation of China and the natural science foundation of Shanxi, China.

References

Beaubouef, T., Petry, F.E. and Arora, G. (1998) "Information-theoretic measures of uncertainty for rough sets and rough relational databases", *Information Sciences* **109**, 185–195.

De Luca, A. and Termini, S. (1972) "A definition of a nonprobabilistic entropy in the setting of fuzzy theory", *Information and Control* **20**, 301–312.

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

Dubois, D. and Prade, H. (1990) "Rough fuzzy sets and fuzzy rough sets", *International Journal of General Systerms* **17**, 191–209.

Hartley, R.V.L. (1928) "Transmission of information", *The Bell Systems Technical Journal* **7**(3), 535–563.

Klir, G.J. and Wierman, M.J. (1998) Uncertainty Based Information (Physica-Verlag, New York).

Liang, J.Y. and Xu, Z.B. (2000) "Uncertainty measures of roughness of knowledge and rough sets in incomplete information systems", Proceedings of the 3rd World Congress on Intelligent Control and Automation (Press of University of Science and Technology of China, Hefei), pp 2526–2529.

Liang, J.Y., Xu, Z.B. and Miao, D.Q. (2000) "Reduction of knowledge in incomplete information systems", Proceedings of Conference on Intelligent Information Processing in 16th World Computer Congress (Publishing House of Electronics Industry, Beijing), pp 528–532.

Lingras, P.J. and Yao, Y.Y. (1998) "Data mining using extensions of the rough set model", *Journal of American Society for Information Science* **49**(5), 415–422.
Liu, X.C. (1992) "Entropy, distance measure and similarity measure of fuzzy sets and their relations", *Fuzzy Sets and Systems* **52**, 305–318.
Pawlak, Z. (1991) Rough Sets: Theoretical Aspects of Reasoning about Data (Kluwer Academic, Dordrecht).
Pawlak, Z., Grzymala-Busse, J.W., Slowiński, R. and Ziarko, W. (1995) "Rough sets", *Communication of ACM* **38**, 89–95.
Shannon, C.E. (1948) "The mathematical theory of communication", *The Bell System Technical Journal* **27**(3 and 4), 373–423 (See also pp. 623–656).
Slowiński, R. (1992) Intelligent Decision Support, Handbook of Applications and Advances of Rough Sets Theory (Kluwer Academic, Dordrecht).
Slowiński, R. and Vanderpooten, D. (2000) "A generalized definition of rough approximations based on similarity", *IEEE Transactions on Knowledge and Data Engineering* **12**(2), 331–336.
Wierman, M.J. (1999) "Measuring uncertainty in rough set theory", *International Journal of General Systems* **28**(4), 283–297.

Jiye Liang is a Professor of Computer Science at Shanxi University, China. Dr Liang received his Ph.D in Information Science from Xi'an Jiaotong University. He also has a B.S. in computational mathematics from Xi'an Jiaotong University. Dr Liang's research interests include the rough set theory, data mining and artificial intelligent.

Dr K. S. Chin is an Associate Professor in the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong. Before joining the University, Dr Chin had more than 10 years of experience in the manufacturing industry. He is a Chartered Engineer in the UK and a Registered Professional Engineer in Hong Kong. Dr Chin is a fellow of Hong Kong Society for Quality, senior members of Institute of Industrial Engineers (IIE), American Society for Quality (ASQ) and Society of Manufacturing Engineers (SME), USA. Dr Chin has published over 80 international refereed papers in the field of engineering management. His current research interests are information management in global supply chain and applications of Artificial Intelligence technologies in engineering management.

Dr Chuangyin Dang received his Ph.D degree (Cum Laude) from Tilburg University in 1991. Currently, he is an Associate Professor of Department of Manufacturing Engineering and Engineering Management. Before joining City University of Hong Kong in 1998, he had taught at the University of California and University of Auckland. Dr Dang's research is in operations research and optimization. He is best known for developing the D_1-triangulation of the Euclidean space and simplicial approaches to integer programming.

Richard C. M. Yam received his Master of Science degree in Management Science from the University of London, Imperial College of Science and Technology and the Ph.D. degree from the University of Warwick, UK. He is now an Associate Professor and a Program Leader of the M.Sc. Engineering Management and the Engineering Doctorate Programmes at City University of Hong Kong. Dr Yam's current research interests include competitiveness analysis, engineering and technology management, and smart asset management.