Applied Mathematics & Information Sciences
An International Journal

Knowledge Entropy and Feature Selection in Incomplete Decision Systems

Jiucheng Xu [1,2] *and Lin Sun* [1,2,3,*]

[1]College of Computer and Information Engineering, Henan Normal University, 453007 Xinxiang, P. R. China
[2]Engineering and Technology Research Center for Computational Intelligence and Data Mining of Universities of Henan Province, 453007 Xinxiang, P. R. China
[3]International WIC Institute, Beijing University of Technology, 100124 Beijing, P. R. China

Received: 27 Jul. 2012; Revised 15 Oct. 2012; Accepted 25 Oct. 2012
Published online: 1 Mar. 2013

Abstract: In this paper, concepts of knowledge entropy and knowledge entropy-based uncertainty measures are given in incomplete information systems and decision systems, and some important properties of them are investigated. From these properties, it can be shown that these measures provide important approaches to measure the uncertainty ability of different knowledge in incomplete decision systems. Then the relationships among these knowledge entropies proposed are discussed as well. A new definition of reduct is proposed and a heuristic algorithm with low computational complexity is constructed to improve computational efficiency of feature selection in incomplete decision systems. Experimental results demonstrate that our algorithm can provide an efficient solution to find a minimal subset of the features from incomplete data sets.

Keywords: Rough set, feature selection, incomplete decision systems, knowledge entropy, uncertainty measure

1. Introduction

Due to the abundance of noisy, irrelevant or misleading features in real world problems, handling imprecise and inconsistent information for feature selection has become one of the most important requirements [1]. In the past two decades, successful applications of rough set model as an effective method to feature selection in a variety of problems have demonstrated its importance and versatility. Starzyk et al. [2] used strong equivalence to simplify discernibility functions. Obviously, this is an expensive solution to the problem and is only practical for very simple data sets. Wroblewski used genetic algorithms to find minimal reducts [3]. However, Wroblewski's method uses time-consuming operations and cannot assure that the resulting subset is really a reduct. Hu et al. [4] used the concept of fuzzy equivalence relation matrix to compute entropy and mutual information for feature selection of real-valued data sets, but at the expense of increased computational effort. In general, these feature selection methods above select the relevant features of a data set without considering the redundancy among them.

It is known that classical rough set theory is unsuitable for feature selection in incomplete decision systems [5]. To address this issue, several interesting and meaningful extensions to equivalence relation have been proposed, such as tolerance relations, neighborhood operators, others [6]. To accomplish feature selection from incomplete decision systems, the heuristic approaches can avoid the exponential computation in exhaustive methods, but they still suffer from intensive computation of tolerance classes induced by the condition features in an incomplete decision system. This process largely affects computational time of feature selection. Many researchers have attempted to solve this problem, however, their time complexities are no less than $O(|C|^2|U|^2)$, where $|C|$ and $|U|$ respectively denote the numbers of condition features and objects. Thus, it is desirable to propose an efficient and effective approach to feature selection in incomplete decision systems.

Heuristic search depends on the measures associated with the features, by which we can analyze the significance of every feature, and regard it as heuristic information in

* Corresponding author: e-mail: linsunok@gmail.com

order to decrease search space. To evaluate uncertainty of a system, the concept of entropy was introduced by Shannon in [7]. The entropy and its variants are adapted for rough set theory in [8,9]. Qian and Liang [10] proposed the concepts of combination entropy and combination granulation to measure the uncertainty of knowledge from a complete information system. These measures include granulation measure, information entropy, rough entropy and knowledge granulation, and have become effective mechanisms for evaluating uncertainty in rough set theory. All these studies are dedicated to evaluating uncertainty of a set in terms of the partition ability of knowledge. However, since the equivalence classes are only regarded as the unit of information granule of a complete information system, these measures cannot be used to deal with an incomplete information system. Therefore, it is desirable to extend and hybridize these measures to deal with incomplete data and solve many real world problems. In this paper, we introduce concepts of some knowledge entropies and uncertainty measures in incomplete information systems and decision systems, and discuss some important properties of them. Furthermore, these measures can provide important approaches to measure the uncertainty ability of different knowledge in incomplete decision systems, and then a heuristic feature selection algorithm with complexity analysis is presented in incomplete decision systems.

2. Preliminaries

The following recalls necessary preliminaries that are relevant to this paper. Detailed description of the theory can be found in the source papers [11,12].

Formally, an information system (IS) is an ordered triple $IS=(U,A,F)$, where U is a nonempty finite set of objects called the universe, and A is a nonempty finite set of features, such that there exists a map $f_a\colon U\to V_a$ for any $a\in A$, where V_a is called the domain of the feature a, and then $F=\{f_a|a\in A\}$. With any $P\subseteq A$, an associated indistinguishable relation is denoted by $IND(P)=\{(u,v)\in U\times U|\forall a\in P, f_a(u)=f_a(v)\}$. The partition of U induced by $IND(P)$ is denoted by $U/IND(P)=\{[u]_P|u\in U\}$, where $[u]_P=\{v\in U|(u,v)\in IND(P)\}$. $U/IND(P)$ can be replaced by U/P.

It may happen that some of the feature values for an object are missing. To indicate such a situation, a distinguished value, the so-called null value is usually assigned to those features. If V_a contains a null value for at least one feature $a\in A$, then IS is called an incomplete information system (IIS), otherwise it is a complete information system (CIS). If $A=C\cup D$ and $C\cap D=\emptyset$, where C is the condition feature set and D is the decision feature set, then IIS is called an incomplete decision system (IDS), and CIS is called a complete decision system (CDS). Further on, the symbol $*$ denotes the missing value. If the value of a feature a is missing, then the real value must be from the set $V_a-\{*\}$. For any $P\subseteq A$, P determines a binary relation $SIM(P)=\{(u,v)\in U\times U|\forall a\in P, f_a(u)=f_a(v)$ or $f_a(u)=*$ or $f_a(v)=*\}$. In fact, $SIM(P)$ is a tolerance relation on U. $S_P(u)=\{v\in U|(u,v)\in SIM(P)\}$ is the maximal set of objects which are possibly indistinguishable by P with u. Let $U/SIM(P)$ denote the family sets $\{S_P(u)|u\in U\}$, the classification or the knowledge induced by P. A member $S_P(u)$ from $U/SIM(P)$ will be called a tolerance class. If $U/SIM(P)=\omega=\{S_P(u)=\{u\}|u\in U\}$, it is called an identity relation, and if $U/SIM(P)=\delta=\{S_P(u)=\{U\}|u\in U\}$, it is called a universal relation.

Let IIS be an incomplete information system. Q is coarser than P, denoted by $P\preceq Q$, if and only if $S_P(u_i)\subseteq S_Q(u_i)$ for $i\in\{1,2,\cdots,|U|\}$. If $P\preceq Q$ and $P\neq Q$, then we say that Q is strictly coarser than P and denoted by $P\prec Q$. In fact, $P\prec Q\Leftrightarrow$ for $i\in\{1,2,\cdots,|U|\}$, it follows that $S_P(u_i)\subseteq S_Q(u_i)$, and there exists $j\in\{1,2,\cdots,|U|\}$ such that $S_P(u_j)\subset S_Q(u_j)$.

3. Knowledge entropy and feature selection

3.1. Incomplete rough entropy and incomplete information entropy of knowledge

Let CIS be a complete information system and $U/A=\{R_1,R_2,\cdots,R_m\}$. Rough entropy of knowledge A is denoted by $E_r(A)=-\sum_{i=1}^{m}\frac{|R_i|}{|U|}\log_2\frac{1}{|R_i|}$, and information entropy of knowledge A is denoted by $H(A)=-\sum_{i=1}^{m}\frac{|R_i|}{|U|}\log_2\frac{|R_i|}{|U|}$.

Property 3.1. Let CIS be a complete information system and $P,Q\subseteq A$. If $P\preceq Q$, then $E_r(P)\le E_r(Q)$ and $H(P)\ge H(Q)$.

When we do not need to distinguish complete information systems and incomplete information systems, an information in IS can be represented as the vector $K(A)=\{S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|})\}$ [13]. Let $U/A=\{X_1, X_2, \cdots, X_m\}$ and $U/SIM(A)=\{S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|})\}$. If 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 the relationship between $K(A)$ and U/A can be established: $X_i=S_A(u_{i1})=S_A(u_{i2})=\cdots=S_A(u_{is_i})$, $|X_i|=|S_A(u_{i1})|=|S_A(u_{i2})|=\cdots=|S_A(u_{is_i})|$, and $|X_i|^2=\sum_{k=1}^{s_i}|S_A(u_{ik})|$ for $i\in\{1,2,\cdots,m\}$. We can obtain that

$$\begin{aligned}E_r(A)&=-\sum_{i=1}^{m}\Big(\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{i1})|}+\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{i2})|}\\&\quad+\cdots+\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{is_i})|}\Big)\\&=-\sum_{i=1}^{m}\sum_{k=1}^{S_i}\frac{1}{|U|}\log_2\frac{1}{|S_A(u_{ik})|}\\&=-\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_A(u_i)|}.\end{aligned}\qquad(1)$$

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

Definition 3.1. Let IIS be an incomplete information system, $P \subseteq A$, $U/SIM(P) = \{S_P(u_1), S_P(u_2), \cdots, S_P(u_{|U|})\}$. Incomplete rough entropy of knowledge P is defined as

$$E(P) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_i)|}. \quad (2)$$

Definition 3.2. Let IIS be an incomplete information system and $P \subseteq A$. Incomplete information entropy of knowledge P is defined as

$$H'(P) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_P(u_i)|}{|U|}. \quad (3)$$

Then, from Definitions 3.1 and 3.2, the following properties can be obtained.

Property 3.2. If $U/SIM(P) = \omega$, then $E(P)$ achieves its minimum value 0, and $H'(P)$ achieves its maximum value $\log_2|U|$. If $U/SIM(P) = \delta$, then $E(P)$ achieves its maximum value $\log_2|U|$, and $H'(P)$ achieves its minimum value 0.

Property 3.3. Let IIS be an incomplete information system and $P, Q \subseteq A$. If there exists a one-to-one, onto function $h : U/SIM(P) \to U/SIM(Q)$ such that $|h(S_P(u_i))| = |S_P(u_i)|$ for $i \in \{1, 2, \cdots, |U|\}$, then $E(P) = E(Q)$ or $H'(P) = H'(Q)$.

Property 3.4. Let IIS be an incomplete information system and $P, Q \subseteq A$. If $P \prec Q$, then $E(P) < E(Q)$.

Proof. Since $P \prec Q$, it follows that $S_P(u_i) \subseteq S_Q(u_i)$ for $i \in \{1, 2, \cdots, |U|\}$, then $|S_P(u_i)| \leq |S_Q(u_i)|$. 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, we can obtain that

$$\begin{aligned} E(P) &= \sum_{i=1, u_i \neq u_0}^{|U|} \frac{1}{|U|} \log_2 |S_P(u_i)| + \frac{1}{|U|} \log_2 |S_P(u_0)| \\ &< \sum_{i=1, u_i \neq u_0}^{|U|} \frac{1}{|U|} \log_2 |S_Q(u_i)| + \frac{1}{|U|} \log_2 |S_Q(u_0)| \\ &= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_Q(u_i)|} \\ &= E(Q). \end{aligned} \quad (4)$$

Property 3.5. Let IIS be an incomplete information system and $P, Q \subseteq A$. If $P \prec Q$, then $H'(P) > H'(Q)$.

Proof. Similar to Property 3.4,

$$\begin{aligned} &- \sum_{i=1, u_i \neq u_0}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_P(u_i)|}{|U|} - \frac{1}{|U|} \log_2 \frac{|S_P(u_0)|}{|U|} \\ &> - \sum_{i=1, u_i \neq u_0}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_Q(u_i)|}{|U|} - \frac{1}{|U|} \log_2 \frac{|S_Q(u_0)|}{|U|}, \end{aligned} \quad (5)$$

i.e., $-\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_P(u_i)|}{|U|} > -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_Q(u_i)|}{|U|}$.

Thus, $H'(P) > H'(Q)$. This completes the proof.

From Properties 3.4 and 3.5, the following property can be obtained immediately.

Property 3.6. Let IIS be an incomplete information system, $P \subseteq A$. $U/SIM(P) = U/SIM(A)$ if and only if $E(P) = E(A)$ or $H'(P) = H'(A)$.

For any $P \subseteq A$ and $r \in P$, we say that r is dispensable if $U/SIM(P) = U/SIM(P - \{r\})$, otherwise it is indispensable in P. P is independent if each $r \in P$ is indispensable in P, otherwise P is dependent. Thus, the following proposition can be obtained immediately.

Proposition 3.1. Let IIS be an incomplete information system. A set $P \subseteq A$ is a reduct of A if and only if P is independent and $E(P) = E(A)$ or $H'(P) = H'(A)$.

Thus, it is concluded that the reduct definition based on tolerance relation and the reduct definition based on incomplete rough (or information) entropy are equivalent in incomplete information systems.

Example 3.1. Consider the descriptions of several cars employed in Table 1 [13], in which $U = \{u_1, u_2, \cdots, u_6\}$, $A = \{P, M, S, X\}$, and P, M, S, X stand for *Price*, *Mileage*, *Size*, and $Max - Speed$.
Assumed that $B = \{P, S, X\}$, it follows that $U/SIM(A) = \{S_A(u_1), S_A(u_2), S_A(u_3), S_A(u_4), S_A(u_5), S_A(u_6)\} = \{\{u_1\}, \{u_2, u_6\}, \{u_3\}, \{u_4, u_5\}, \{u_4, u_5, u_6\}, \{u_2, u_5, u_6\}\}$, and $U/SIM(B) = \{S_B(u_1), S_B(u_2), S_B(u_3), S_B(u_4), S_B(u_5), S_B(u_6)\} = \{\{u_1\}, \{u_2, u_6\}, \{u_3\}, \{u_4, u_5\}, \{u_4, u_5, u_6\}, \{u_2, u_5, u_6\}\}$. It is easily computed that $U/SIM(B) = U/SIM(A)$, and $E(B) = E(A) = 0.862$. Furthermore, the others are not equal to $U/SIM(A)$. Therefore, $\{P, S, X\}$ is a reduct of $\{P, M, S, X\}$.

Proposition 3.2. Let IIS be an incomplete information system and $P \subseteq A$. Then $E(P) + H'(P) = \log_2|U|$.

Proof. It follows from Definitions 3.1 and 3.2 that

$$\begin{aligned} &E(P) + H'(P) \\ &= -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{|S_P(u_i)|} - \sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_P(u_i)|}{|U|} \\ &= -\frac{1}{|U|} \sum_{i=1}^{|U|} (\log_2 \frac{1}{|S_P(u_i)|} + \log_2 \frac{|S_P(u_i)|}{|U|}) \\ &= \log_2 |U|. \end{aligned} \quad (6)$$

In what follows, we investigate the incomplete rough (or information) entropy of new knowledge composed of two given knowledge with the same universe in incomplete information systems.

Lemma 3.1. Let IIS be an incomplete information system and $P, Q \subseteq A$. $SIM(P) \cap SIM(Q) = SIM(P \cup Q)$ and $S_P(u) \cap S_Q(u) = SIM_{P \cup Q}(u)$ hold.

Lemma 3.2. Let IIS be an incomplete information system and $P, Q \subseteq A$. Then $U/SIM(P) \cap U/SIM(Q) = U/SIM(P \cup Q)$.

Definition 3.3. Let IIS be an incomplete information system and $P,Q \subseteq A$. Incomplete rough entropy of knowledge $P\cup Q$ is defined as

$$\begin{aligned} E(P\cup Q) &= -\sum_{i=1}^{|U|}\sum_{j=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)\cap S_Q(u_j)|} \\ &= -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)\cap S_Q(u_i)|}. \end{aligned} \tag{7}$$

Definition 3.4. Let IIS be an incomplete information system and $P,Q \subseteq A$. Incomplete information entropy of knowledge $P\cup Q$ is defined as

$$\begin{aligned} H'(P\cup Q) &= -\sum_{i=1}^{|U|}\sum_{j=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)\cap S_Q(u_j)|}{|U|} \\ &= -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)\cap S_Q(u_i)|}{|U|}. \end{aligned} \tag{8}$$

From Definitions 3.3 and 3.4, the following properties can be obtained.

Proposition 3.3. Let IIS be an incomplete information system and $P,Q \subseteq A$. Then the following properties hold
(1) $E(P\cup Q) \le E(P)$ and $E(P\cup Q) \le E(Q)$.
(2) $H'(P\cup Q) \ge H'(P)$ and $H'(P\cup Q) \ge H'(Q)$.
(3) If $P \prec Q$, then $E(P\cup Q) = E(Q)$, and $H'(P\cup Q) = H'(Q)$.
(4) $E(P\cup Q) + H'(P\cup Q) = \log_2|U|$.

3.2. Incomplete conditional entropy of knowledge in incomplete decision systems

Definition 3.5. Let IDS be an incomplete decision system, $P \subseteq C$. Incomplete conditional rough entropy of D with reference to P is defined as

$$E(D|P) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|S_P(u_i)\cap S_D(u_i)|}{|S_P(u_i)|}. \tag{9}$$

Definition 3.6. Let IDS be an incomplete decision system, $P \subseteq C$. Incomplete conditional information entropy of D with reference to P is defined as

$$H'(D|P) = -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)\cap S_D(u_i)|}{|S_P(u_i)|}. \tag{10}$$

Then, from Definitions 3.5 and 3.6, the following property can be obtained immediately.

Property 3.7. Let IDS be an incomplete decision system, $P \subseteq C$. Then $E(D|P) = H'(D|P)$.

Proposition 3.4. Let IDS be an incomplete decision system, $P \subseteq C$. Then $E(D|P) = E(P) - E(P\cup D)$ and $H'(D|P) = H'(P\cup D) - H'(P)$.

Proof. From Definitions 3.1, 3.3 and 3.5, we have

$$\begin{aligned} &E(D|P) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}(\log_2|S_P(u_i)\cap S_D(u_i)| - \log_2|S_P(u_i)|) \\ &= -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)|} + \sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_i)\cap S_D(u_i)|} \\ &= E(P) - E(P\cup D). \end{aligned} \tag{11}$$

Similarly, from Definitions 3.2, 3.4 and 3.6, we obtain

$$\begin{aligned} &H'(D|P) \\ &= -\sum_{i=1}^{|U|}\frac{1}{|U|}(\log_2\frac{|S_P(u_i)\cap S_D(u_i)|}{|U|} - \log_2\frac{|S_P(u_i)|}{|U|}) \\ &= -\sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)\cap S_D(u_i)|}{|U|} + \sum_{i=1}^{|U|}\frac{1}{|U|}\log_2\frac{|S_P(u_i)|}{|U|} \\ &= H'(P\cup D) - H'(P). \end{aligned} \tag{12}$$

Definition 3.7. Let IDS be an incomplete decision system, $P \subseteq C$. Mutual information between P and D is defined as

$$E(P;D) = E(D) - E(P) + E(P\cup D), \tag{13}$$

$$H'(P;D) = H'(P) + H'(D) - H'(P\cup D). \tag{14}$$

Note that the relationships among information entropy, conditional information entropy and mutual information can satisfy an identical equation in a complete information system [7,9]. However, so far the above relationships have not been reported in incomplete decision systems, thus, we further investigate the relationships among these measures above as follows.

Proposition 3.5. Let IDS be an incomplete decision system and $P \subseteq C$. Then
(1) $E(P;D) = E(D) - E(D|P)$.
(2) $H'(P;D) = H'(D) - H'(D|P)$.
(3) $E(D) - E(P;D) = H'(D) - H'(P;D)$.

Proof. (1) From Definition 3.7 and Proposition 3.4, we have that $E(P;D) = E(D) - E(P) + E(P\cup D) = E(D) - (E(P) - E(P\cup D)) = E(D) - E(D|P)$.
(2) Similar to (1), the equation $H'(P;D) = H'(D) - H'(D|P)$ can be proved.
(3) From Property 3.7, (1) and (2), $E(D) - E(P;D) = H'(D) - H'(P;D)$ is straightforward.

Remark. Propositions 3.4 and 3.5 establish the relationships among the incomplete rough entropy, the incomplete conditional rough entropy, the incomplete information entropy, the incomplete conditional information entropy and the mutual information in incomplete decision systems. In fact, these equations cannot be satisfied by some existing measures in incomplete information systems and incomplete decision systems. These relationships will be helpful for understanding the essence of the knowledge content and the uncertainty in an incomplete decision system. Several authors have applied Shannon's entropy and its variants for measuring the uncertainty in the context of complete information systems [8]. However, these entropies cannot be used in incomplete information systems. Thus, to overcome the shortcomings of some existing uncertainty measures, the knowledge entropies above can measure the uncertainty in incomplete information systems and incomplete decision systems. Note that the incomplete conditional rough (or information) entropy can measure both the uncertainty in incomplete decision systems and that in complete decision systems. To this end, we investigate some properties of conditional rough (or information) entropy in complete decision systems. It is known that equivalence relation is a kind of special tolerance relation [5]. Thus, some relevant definitions and propositions of conditional entropy in complete decision systems are proposed as follows.

Definition 3.8. Let *CDS* be a complete decision system, $P \subseteq C$, $U/P = \{X_1, X_2, \cdots, X_m\}$, and $U/D = \{Y_1, Y_2, \cdots, Y_n\}$. Conditional rough entropy of D with reference to P on U is defined as

$$E_r(D|P) = \sum_{i=1}^{m} \frac{|X_i|}{|U|} \sum_{j=1}^{n} \frac{|Y_j \cap X_i|}{|X_i|} \log_2 \frac{|X_i|}{|Y_j \cap X_i|}. \quad (15)$$

For any $P \subseteq C$, conditional information entropy of D with reference to P on U is denoted by

$$H(D|P) = -\sum_{i=1}^{m} \frac{|X_i|}{|U|} \sum_{j=1}^{n} \frac{|Y_j \cap X_i|}{|X_i|} \log_2 \frac{|Y_j \cap X_i|}{|X_i|}. \quad (16)$$

Thus, one has that $H(D|P) = H(P \cup D) - H(P)$ [9].

Property 3.8. Let *CDS* be a complete decision system, $P \subseteq C$. Then $E_r(D|P) = H(D|P)$ and $E_r(D|P) = E_r(P) - E_r(P \cup D)$.

Proposition 3.6. Let *CDS* be a complete decision system. $U/A_1 = \{X_1, X_2, \cdots, X_n\}$ is a partition of U induced by A_1. $U/A_2 = \{X_1, X_2, \cdots, X_{p-1}, X_{p+1}, \cdots, X_{q-1}, X_{q+1}, \cdots, X_n, X_p \cup X_q\}$ is another partition generated through combining equivalence blocks X_p and X_q to $X_p \cup X_q$. Then we have that $E_r(D|A_1) \leq E_r(D|A_2)$.

Proof. The proof is similar to Lemma 4.1 [9].

Corollary 3.1. Let *CDS* be a complete decision system and $A_1, A_2 \subseteq C$. Then $E_r(D|A_1) \leq E_r(D|A_2)$ if and only if $H(D|A_1) \leq H(D|A_2)$.

Corollary 3.2. Let *CDS* be a complete decision system and $P \subset Q \subseteq C$. Then $E_r(D|Q) < E_r(D|P)$ and $H(D|Q) < H(D|P)$.

Proposition 3.7. Let *IDS* be an incomplete decision system and $P \subset Q \subseteq C$. Then $E(D|Q) < E(D|P)$ does not always hold.

Proof. For any $P \subset Q$, then $Q \prec P$. It follows that $S_Q(u_i) \subseteq S_P(u_i)$ for $i \in \{1, 2, \cdots, |U|\}$, and there exists $j \in \{1, 2, \cdots, |U|\}$ such that $S_Q(u_j) \subset S_P(u_j)$. Then, suppose $i \in \{1, 2, \cdots, t\}$, there exists $S_Q(u_i) \subseteq S_P(u_i)$ such that for $j \in \{t+1, t+2, \cdots, |U|\}$, $S_Q(u_j) \subset S_P(u_j)$, where $0 \leq t < |U|$. Case1: Suppose $i \in \{1, 2, \cdots, t\}$, it follows that $S_Q(u_i) \subseteq S_P(u_i)$ and $S_D(u_i) \cap S_Q(u_i) \subseteq S_D(u_i) \cap S_P(u_i)$. That is, $|S_Q(u_i)| \leq |S_P(u_i)|$ and $|S_D(u_i) \cap S_Q(u_i)| \leq |S_D(u_i) \cap S_P(u_i)|$. Thus, $\frac{|S_Q(u_i)|}{|S_P(u_i)|} \leq 1$ and $\frac{|S_P(u_i) \cap S_D(u_i)|}{|S_Q(u_i) \cap S_D(u_i)|} \geq 1$ for $i \in \{1, 2, \cdots, t\}$. Obviously, there exists only $S_Q(u_i) = S_P(u_i)$ for $i \in \{1, 2, \cdots, t\}$, then the equality above holds. Case2: Suppose $j \in \{t+1, t+2, \cdots, |U|\}$, then $S_Q(u_j) \subset S_P(u_j)$ and $S_D(u_j) \cap S_Q(u_j) \subset S_D(u_j) \cap S_P(u_j)$. That is, $|S_Q(u_j)| < |S_P(u_j)|$ and $|S_D(u_j) \cap S_Q(u_j)| < |S_D(u_j) \cap S_P(u_j)|$. Hence, we have that

$$\begin{aligned} &E(D|Q) - E(D|P) \\ &= \sum_{i=1}^{t} \frac{1}{|U|} \log_2 \frac{|S_Q(u_i)||S_P(u_i) \cap S_D(u_i)|}{|S_P(u_i)||S_Q(u_i) \cap S_D(u_i)|} \\ &+ \sum_{j=t+1}^{|U|} \frac{1}{|U|} \log_2 \frac{|S_Q(u_j)||S_P(u_j) \cap S_D(u_j)|}{|S_P(u_j)||S_Q(u_j) \cap S_D(u_j)|}. \end{aligned} \quad (17)$$

Therefore, it follows from $Q \prec P$ that there exist $\frac{|S_Q(u_k)|}{|S_P(u_k)|} < 1$ and $\frac{|S_P(u_k) \cap S_D(u_k)|}{|S_Q(u_k) \cap S_D(u_k)|} > 1$ for any $u_k \in U$ such that $E(D|Q) > E(D|P)$ or $E(D|Q) < E(D|P)$. Then, $E(D|Q) < E(D|P)$ does not always hold.

Remark. Proposition 3.7 states that incomplete conditional rough entropy of knowledge does not decrease monotonically with increases of features in knowledge through finer classification, which means that adding a new feature into the existing subset of condition features will change the value of incomplete conditional rough entropy.

Corollary 3.3. Let *IDS* be an incomplete decision system and $P \subset Q \subseteq C$. Then $H'(D|Q) < H'(D|P)$ does not always hold.

Proposition 3.8. Let *CDS* be a complete decision system, $P \subseteq C$. Then $E(D|P) = E_r(D|P) = H(D|P) = H'(D|P)$.

Proof. Let $X_i \in U/P$, $X_i = \{u_{i1}, u_{i2}, \cdots, u_{is_i}\}$, $|X_i| = s_i$, and $\sum_{i=1}^{m} s_i = |U|$. The relationships among the elements in $U/SIM(P)$ and the elements in U/P are as follows: $X_i = S_P(u_{i1}) = S_P(u_{i2}) = \cdots = S_P(u_{is_i})$, i.e., $|X_i| = |S_P(u_{i1})| = |S_P(u_{i2})| = \cdots = |S_P(u_{is_i})|$. Similarly, let $Y_j \in U/D$, $|Y_j| = \{u_{j1}, u_{j2}, \cdots, u_{jt_j}\}$, and $\sum_{j=1}^{n} t_j = |U|$. The relationships among the elements in $U/SIM(D)$ and the elements in U/D are as follows: $Y_j = S_D(u_{j1}) = S_D(u_{j2}) = \cdots = S_D(u_{jt_j})$. Thus, we can obtain that

$$\begin{aligned} E(D|P) &= \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{u_k \in Y_j \cap X_i} \frac{1}{|U|} \log_2 \frac{|S_P(u_k)|}{|S_P(u_k) \cap S_D(u_k)|} \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{|Y_j \cap X_i|}{|U|} \log_2 \frac{|X_i|}{|Y_j \cap X_i|} \\ &= \sum_{i=1}^{m} \frac{|X_i|}{|U|} \sum_{j=1}^{n} \frac{|Y_j \cap X_i|}{|X_i|} \log_2 \frac{|X_i|}{|Y_j \cap X_i|} \\ &= E_r(D|P). \end{aligned} \quad (18)$$

From Properties 3.7 and 3.8, $H'(D|P) = E(D|P) = E_r(D|P) = H(D|P)$ hold. This completes the proof.

Proposition 3.8 states that the incomplete conditional rough (or information) entropy under tolerance relation is the extended formulation of the complete conditional rough (or information) entropy under equivalence relation.

Proposition 3.9. Let CIS be a complete information system, $P \subseteq A$. Then $E(P) = E_r(P)$ and $H'(P) = H(P)$.

Proof. Similar to Proposition 3.8,

$$\begin{aligned} E(P) &= -\sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\log_2\frac{1}{|S_P(u_{ik})|} \\ &= -\sum_{i=1}^{m}\frac{|X_i|}{|U|}\log_2\frac{1}{|X_i|} \\ &= E_r(P). \end{aligned} \tag{19}$$

Similarly, the equation $H'(P) = H(P)$ can be proved. This completes the proof.

Propositions 3.8 and 3.9 state that the four uncertainty measures above, named as the incomplete conditional rough (or information) entropy and the incomplete rough (or information) entropy, are unified in both incomplete and complete decision systems and both incomplete and complete information systems respectively. Unlike all existing measures for the uncertainty in incomplete information systems, incomplete and complete decision systems, the relationships among them can be established, which are formally expressed as follows: (a) $E(P) + H'(P) = \log_2 |U|$ in incomplete information systems; (b) $E(P) = E_r(P)$ and $H'(P) = H(P)$ in complete information systems; (c) $E(D|P) = E(P) - E(P \cup D)$, $H'(D|P) = H'(P \cup D) - H'(P)$, $E(D|P) = H'(D|P)$, $E(P;D) = E(D) - E(D|P)$, $H'(P;D) = H'(D) - H'(D|P)$, and $E(D) - E(P;D) = H'(D) - H'(P;D)$ in incomplete decision systems; (d) $E_r(D|P) = E_r(P) - E_r(P \cup D)$ and $E(D|P) = E_r(D|P) = H(D|P) = H'(D|P)$ in complete decision systems. These relationships are very significant for reasonably applying an uncertainty measure to incomplete information systems and incomplete decision systems. So far, however, the relationships above have not been reported in incomplete decision systems, which would hinder further research and application of knowledge entropy. Furthermore, all existing knowledge entropies and their extensions in incomplete information systems and in incomplete decision systems cannot establish the above relationships. Therefore, the incomplete conditional rough (or information) entropy may be a much better uncertainty measure to evaluate the knowledge of an incomplete decision system.

3.3. Incomplete conditional entropy-based reduct in incomplete decision systems

From Property 3.7, the incomplete conditional rough (or information) entropy is named as the incomplete conditional entropy in short. Let CDS be a complete decision system and $P \subseteq C$. Then,

$$\begin{aligned} E(D|P) &= \sum_{i=1}^{m}\Big(\frac{1}{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_{i1})|}{|S_P(u_{i1})|}\log_2\frac{|S_P(u_{i1})|}{|Y_j \cap S_P(u_{i1})|} \\ &\quad + \frac{1}{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_{i2})|}{|S_P(u_{i2})|}\log_2\frac{|S_P(u_{i2})|}{|Y_j \cap S_P(u_{i2})|} + \cdots \\ &\quad + \frac{1}{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_{is_i})|}{|S_P(u_{is_i})|}\log_2\frac{|S_P(u_{is_i})|}{|Y_j \cap S_P(u_{is_i})|}\Big) \\ &= \sum_{i=1}^{m}\sum_{k=1}^{s_i}\frac{1}{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_{ik})|}{|S_P(u_{ik})|}\log_2\frac{|S_P(u_{ik})|}{|Y_j \cap S_P(u_{ik})|} \\ &= \frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_i)|}{|S_P(u_i)|}\log_2\frac{|S_P(u_i)|}{|Y_j \cap S_P(u_i)|}. \end{aligned} \tag{20}$$

Thus, the incomplete conditional entropy of D with reference to P on U in incomplete decision systems is reformulated as follows:

$$CE(D|P) = \frac{1}{|U|}\sum_{i=1}^{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_i)|}{|S_P(u_i)|}\log_2\frac{|S_P(u_i)|}{|Y_j \cap S_P(u_i)|}. \tag{21}$$

Without loss of generality, one can obtain that $U/D = \{Y_1, Y_2, \cdots, Y_n\}$ in an incomplete decision system, then the relationship between U/D and $U/SIM(D)$ can be established as follows: $Y_j = S_D(u_{j1}) = S_D(u_{j2}) = \cdots = S_D(u_{jt_j})$ for $j \in \{1, 2, \cdots, n\}$. Thus, the transformed formula of the incomplete conditional entropy is equivalent to that of Definition 5.

Definition 3.9. Let IDS be an incomplete decision system, $P \subseteq C$. Incomplete conditional entropy of D with reference to P on $u_i \in U$ for $i \in \{1, 2, \cdots, |U|\}$ is defined as

$$CE(D|P)_{u_i} = \frac{1}{|U|}\sum_{j=1}^{n}\frac{|Y_j \cap S_P(u_i)|}{|S_P(u_i)|}\log_2\frac{|S_P(u_i)|}{|Y_j \cap S_P(u_i)|}. \tag{22}$$

Definition 3.10. Let IDS be an incomplete decision system. A set $P \subseteq C$ is called a reduct of the IDS if and only if $CE(D|C)_{u_i} \le CE(D|P)_{u_i}$ for $i \in \{1, 2, \cdots, |U|\}$, and for any $Q \subset P$, there exists $u_j \in U$ such that $CE(D|Q)_{u_j} > CE(D|C)_{u_j}$.

Proposition 3.10. Let IDS be an incomplete decision system, $P \subseteq C$. If P is a reduct of C with reference to D, then $CE(D|C) \le CE(D|P)$ on $u_i \in U$ for $i \in \{1, 2, \cdots, |U|\}$.

Proof. Since $P \subseteq C$ and P is a reduct of C with reference to D, it follows from Definition 3.10 that for any $u_i \in U$, then $CE(D|C)_{u_i} < CE(D|P)_{u_i}$. Thus, one has that $CE(D|C) = \sum_{i=1}^{|U|} CE(D|C)_{u_i} \le \sum_{i=1}^{|U|} CE(D|P)_{u_i} = CE(D|P)$. This completes the proof.

Proposition 3.10 provides an approach to judge monotonicity of the incomplete conditional entropy under reduction. In fact, Propositions 3.8 and 3.10 present the theoretical foundation to construct our reduct algorithm, and provide an effective way of determining whether or not a feature should be contained in a reduct. Moreover, it is concluded that the judgement propositions of reduct above are suitable for both incomplete and complete decision systems.

3.4. Feature selection algorithm

To improve computational efficiency of the heuristic feature selection algorithm in incomplete data sets, we construct the input sequence by sorting the features in ascending order via their incomplete conditional entropy. The optimal reduct can be found by repeatedly deleting the head node (feature) of the input sequence. Then we introduce the idea of radix sorting in [14] to calculate tolerance blocks effectively. In fact, there may be multiple reducts for a given decision system. In most applications, it is enough to find a single optimal reduct. Based on the ideas above, we present a fast heuristic incomplete conditional entropy-based feature selection algorithm (called FSICE) as follows.

Algorithm FSICE
Input: An incomplete decision system *IDS*, in which $C = \{c_1, c_2, \cdots, c_p\}$, $D = \{d\}$, and $p = |C|$
Output: *reduct*, a reduct of the *IDS*
(1) Calculate U/D and $S_C(u_i)$ based on radix sorting to obtain $CE(D|C)_{u_i}$ for any $u_i \in U$, $i \in \{1, 2, \cdots, |U|\}$
(2) Construct an ascending input sequence which is ordered by $CE(D|\{c_i\})$, where $c_i \in C$, $i \in \{1, 2, \cdots, p\}$, and denote the result by $\langle a_1, a_2, \cdots, a_p \rangle$
(3) Let $R = \{a_1, a_2, \cdots, a_p\}$
(4) For ($i = 1$; $i \le p$; $i{+}{+}$)
  (4.1) Let $R = R - \{a_i\}$
  (4.2) Calculate $CE(D|R)_{u_j}$, where $j \in \{1, 2, \cdots, |U|\}$
  (4.3) If $CE(D|C)_{u_j} \le CE(D|R)_{u_j}$ for $j \in \{1, 2, \cdots, |U|\}$, break
(5) Construct a descending input sequence and denote the result by $R = \{a_1, a_2, \cdots, a_t\}$, where $t < p$
(6) For ($i = 1$; $i \le t$; $i{+}{+}$)
  (6.1) Let $Q = R - \{a_i\}$
  (6.2) Calculate $CE(D|Q)_{u_j}$, where $j \in \{1, 2, \cdots, |U|\}$
  (6.3) If $CE(D|Q)_{u_j} \le CE(D|C)_{u_j}$ for $j \in \{1, 2, \cdots, |U|\}$, then $R = Q$
(7) *reduct* = R

We first give a fast computing for acquiring tolerance blocks based on the idea of radix sorting algorithm [14], and its complexity is cut down to $O(|C||U|)$. Then the complexity of computing incomplete conditional entropy is $O(|U|\log|U|)$. Thus, the time complexity of (2) in FSICE is no more than $O(|C||U|\log|U|)$. Here, we construct an input sequence (the complexity is $O(|C||U|)$), so that the corresponding tolerance blocks are easily obtained by scanning only the data sets one time. Then, the complexity from (2) to (5) in FSICE is no more than $O(|C||U|\log|U| + |U|)$, and the complexity of (6) in FSICE is also no more than $O(|R||U|\log|U| + |U|)$. Therefore, obviously, the time complexity of Algorithm FSICE is $O(|C||U|\log|U|) + O(|R||U|\log|U| + |U|)$. For the large data sets, in especial gene expressing data sets, since $|U| << |C|$, the whole time complexity of Algorithm FSICE is $O(|C||U|\log|U|)$, which is less than that in [9,11], and its space complexity is $O(|C||U|)$.

4. Experimental results

In this section, we apply the proposed approach and other feature selection approaches to several data sets from UCI databases, so as to evaluate the performances of our approach on a personal computer equipped with Intel Pentium (R) dual-core CPU 2.8GHz and 2GB Memory.

Example 4.1. Consider an incomplete decision system about several cars employed in Table 1, where $C = \{Price, Mileage, Size, Max-Speed\} = \{P, M, S, X\}$ and $D = \{Acceleration\}$.

Table 1 An incomplete decision system about cars

car	P	M	S	X	D
1	High	Low	Full	Low	Good
2	Low	*	Full	Low	Good
3	*	*	Compact	Low	Poor
4	High	*	Full	High	Good
5	*	*	Full	High	Excellent
6	Low	High	Full	*	Good

(1) By computing, it follows that $U/D = \{Y_1, Y_2, Y_3\} = \{\{1,2,4,6\}, \{3\}, \{5\}\}$, $U/SIM(C) = \{S_C(1), S_C(2), S_C(3), S_C(4), S_C(5), S_C(6)\} = \{\{1\}, \{2,6\}, \{3\}, \{4,5\}, \{4,5,6\}, \{2,5,6\}\}$. Then, $CE(D|C)_{u_1} = CE(D|C)_{u_2} = CE(D|C)_{u_3} = 0$, $CE(D|C)_{u_4} = 0.167$, and $CE(D|C)_{u_5} = CE(D|C)_{u_6} = 0.153$.
(2) $U/SIM(\{P\}) = \{\{1,3,4,5\}, \{2,3,5,6\}, \{1,2,3,4,5,6\}\}$, then $CE(D|\{P\}) = 10.503$. $U/SIM(\{M\}) = \{\{1,2,3,4,5\}, \{1,2,3,4,5,6\}, \{2,3,4,5,6\}\}$, then $CE(D|\{M\}) = 7.614$. $U/SIM(\{S\}) = \{\{1,2,4,5,6\}, \{3\}\}$, then $CE(D|\{S\}) = 6.533$. $U/SIM(\{X\}) = \{\{1,2,3,6\}, \{4,5,6\}\}$, then we have $CE(D|\{X\}) = 6.190$. Hence, the input sequence is $\{X, S, M, P\}$.
(3) Let $R = \{X, S, M, P\}$.
(4) When $i = 1$, $R = R - \{X\} = \{S, M, P\}$, then $CE(D|R)_{u_1} = CE(D|R)_{u_2} = CE(D|R)_{u_4} = CE(D|R)_{u_6} = 0.153$, $CE(D|R)_{u_3}$ =0, and $CE(D|R)_{u_5} = 0.120$. So, we find that $CE(D|C)_{u_4} > CE(D|R)_{u_4}$ and $CE(D|C)_{u_5} > CE(D|R)_{u_5}$.
(5) When $i = 2$, $R = R - \{S\} = \{X, M, P\}$, then $CE(D|R)_{u_1} = CE(D|R)_{u_4} = 0.167$, $CE(D|R)_{u_2} = CE(D|R)_{u_3} = CE(D|R)_{u_5} = 0.153$, and $CE(D|R)_{u_6} = 0.250$. We obtain that $CE(D|C)_{u_j} \le CE(D|R)_{u_j}$, where $j = 1, 2, \cdots, 6$.
(6) Let $R = \{P, M, X\}$.
(7) When $i = 1$, $Q = R - \{P\} = \{M, X\}$, then $CE(D|Q)_{u_1} = CE(D|Q)_{u_2} = CE(D|Q)_{u_3} = CE(D|Q)_{u_4} = CE(D|Q)_{u_5} = 0.153$, and $CE(D|Q)_{u_6} = 0.228$. Thus, we have $CE(D|C)_{u_4} > CE(D|Q)_{u_4}$.

(8) When i = 2, $Q = R - \{M\} = \{P,X\}$, it follows that $CE(D|Q)_{u_1} = CE(D|Q)_{u_4} = 0.167, CE(D|Q)_{u_2} = CE(D|Q)_{u_3} = CE(D|Q)_{u_5} = 0.153$, and $CE(D|Q)_{u_6} = 0.250$. We find that $CE(D|C)_{u_j} \leq CE(D|Q)_{u_j}$, where $j = 1,2,\cdots,6$.
(9) *reduct* = $R = \{X,P\}$.

As a result, $\{X,P\}$ is a minimal reduct of the incomplete decision system, while the searching result of the feature selection algorithms in [5,11] is $\{X,P,S\}$. Therefore, its length of $\{X,P,S\}$ is greater than that of the minimal reduct $\{X,P\}$.

Next is the second part of our experiments. The approach [11] has something of a head start in solving the problem of feature selection for incomplete information systems. Then, we have used principles from ref. [11] to design Alg_a for feature selection in decision systems. Meng and Shi [5] proposed a fast positive region-based approach to feature selection. For convenience, this section refers to Algorithm 7 in [5] as Alg_b. Thus, we compare Algorithm FSICE with Alg_a and Alg_b on eight data sets from UCI database, outlined in Table 2. There are two data sets with missing values: Soybean and Vote. To construct incomplete decision systems, we randomly change some feature values from each data set into missing values. The feature selection results of three algorithms on different data sets are summarized in Table 2, in which the minimal cardinality each algorithm finds is given. Here, we always present the optimal solution. From Table 2, it can be seen that the performances of FSICE and Alg_b are very close, though FSICE performs a little better than Alg_a and Alg_b.

Table 2 Number of selected features for each algorithm

Data sets	Samples	Features	Alg_a	Alg_b	FSICE
Zoo	101	17	6	5	5
Balance	200	5	4	4	4
Led24	200	25	12	12	11
Spect	267	23	15	14	14
Soybean	307	36	11	9	9
Vote	435	17	14	9	8
DNAstalog	2000	61	12	10	10
Mushroom	8142	23	6	5	4

As is suggested by Jensen and Shen [12], each data set is tested 20 times. The running time of each algorithm is the average CPU time, shown in Fig.1. It can be seen from Fig.1 that although both running times of Alg_a and Alg_b increase with size of data sets, the former increases much more rapidly than the latter. This difference can be illustrated by plotting the ratios of their running times (Fig. 2). From Fig. 2, we find that the slope of the two curves, shown by Alg_a/FSICE and Alg_b/FSICE, also tends to increase with size of data sets, and the performances of Alg_a and Alg_b are very close, shown by Alg_a/Alg_b. This relationship is not strictly monotonic, and the curve fluctuates distinctly. In fact, the main reason is that the feature number of data sets is different. For example, the DNAstalog has 61 features while the Balance has only 5. Thus, one could envision that this situation must have occurred because of different numbers of features selected. Therefore, FSICE can provide the best running times, and the results show that the proposed method is more suitable to find minimal reducts from incomplete large data sets.

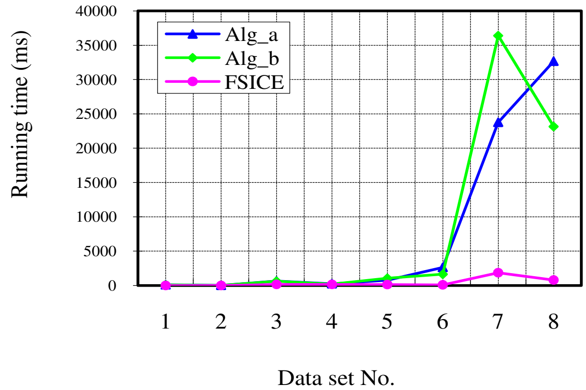

Figure 1 Running time in eight data sets

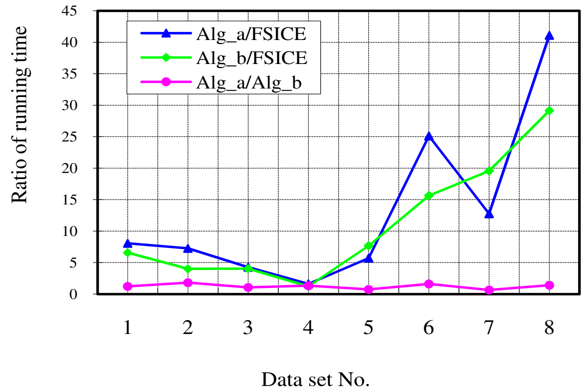

Figure 2 Ratio of the running time in eight data sets

5. Conclusions

Feature selection method, processing the features in the large sample, high-dimension, and complex data set, is an active research field in recent years. Rough sets as a feature selection method can handle imprecision, uncertainty and vagueness [12,13]. Until now, many methods have been proposed to solve the uncertainty measure by using rough set [15,16]; however, each method has some limitation itself and they are unsuitable for incomplete decision systems. So, in this paper, the concepts of knowledge entropies and mutual information are introduced to measure the uncertainty of knowledge in incomplete information systems and decision systems. Then some important properties of these measures are investigated. Furthermore, these measures overcome some disadvantages of classical methods of uncertainty measure, and provide important approaches to measure the uncertainty ability of different knowledge in incomplete decision systems. Moreover, the relationships among the knowledge entropies proposed are established as well. A fast heuristic feature selection algorithm for both incomplete and complete decision systems is developed. Finally, experiments from an example given and eight UCI data sets illustrate that the proposed algorithm is effective and efficient, and the experiment results are consistent with our theoretical analysis.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant (No. 60873104, No. 6104 0037), the Key Project of Science and Technology Department of Henan Province (No. 112102210194), the Science and Technology Research Key Project of Educational Department of Henan Province (No. 12A520027), the Education Fund for Youth Key Teachers of Henan Normal University, and the Postgraduate Science and Technology Program of Beijing University of Technology (No. ykj-2012-6765).

References

[1] X.Y. Wang, J. Yang, X.L. Teng, W.J. Xia and R. Jensen, Feature selection based on rough sets and particle swarm optimization, Pattern Recognition Letters **28**, 459-471 (2007).
[2] J. Starzyk, D.E. Nelson and K. Sturtz, Reduct generation in information systems, Bulletin of International Rough Set Society **3**, 19-22 (1998).
[3] J. Wroblewski, Finding minimal reducts using genetic algorithms, Proceedings of the Second Annual Join Conference on Information Sciences, 186-189 (1995).
[4] Q.H. Hu, D.R. Yu, Z.X. Xie and J.F. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Transactions on Fuzzy Systems **14**, 191-201 (2006).
[5] Z.Q. Meng and Z.Z. Shi, A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets, Information Sciences **179**, 2774-2793 (2009).
[6] W.H. Xu, X.Y. Zhang and W.X. Zhang, Knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems, Applied Soft Computing **9**, 1244-1251 (2009).
[7] C.E. Shannon, The mathematical theory of communication, Bell System Technical Journal **27**, 373-423 (1948).
[8] J.Y. Liang and D.Y. Li, Uncertainty and Knowledge Acquisition in Information Systems, Science Press, Beijing, 2005.
[9] G.Y. Wang, Rough reduction in algebra view and information view, International Journal of Intelligent Systems **18**, 679-688 (2003).
[10] Y.H. Qian and J.Y. Liang, Combination entropy and combination granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **16**, 179-193 (2008).
[11] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences **112**, 39-49 (1998).
[12] L. Sun, J.C. Xu and Y. Tian, Feature selection using rough entropy-based uncertainty measures in incomplete decision systems, Knowledge-Based Systems **36**, 206- 216 (2012).
[13] J.Y. Liang and Y.H. Qian, Information granules and entropy theory in information systems, Science in China Series F: Information Sciences **51**, 1427-1444 (2008).
[14] Z.Y. Xu, Z.P. Liu, B.R. Yang and W. Song, A quick attribute reduction algorithm with complexity of max $(O(|C||U|), O(|C|^2|U/C|))$, Chinese Journal of Computers **29**, 391-399 (2006).
[15] L. Sun, J.C. Xu, J.Y. Ren, T.H. Xu and Q.Q. Zhang, Granularity partition-based feature selection and its application in decision systems, Journal of Information and Computational Science **9**, 3487-3500 (2012).
[16] L. Sun, J.C. Xu, Z.A. Xue and L.J. Zhang, Rough entropy-based feature selection and its application, Journal of Information and Computational Science **8**, 1525-1532 (2011).

Jiucheng Xu is currently a Professor at College of Computer & Information Engineering, Henan Normal University. He received his B.S. degree in Mathematics, Henan Normal University in 1986, the M.S. degree and the Ph.D. degree in Computer Science and Technology, Xi'an Jiaotong University in 1995 and 2004, respectively. His main research interests include granular computing, rough set, data mining, and intelligent information processing.

Lin Sun is currently a Ph.D. Candidate in School of Electronic Information and Control Engineering, Beijing University of Technology. He received his B.S. and M.S. degree in Computer Science and Technology, Henan Normal University in 2003 and 2007, respectively. His main research interests include granular computing, rough set, and data mining.