

## Research Article

# A Novel Approach of Rough Conditional Entropy-Based Attribute Selection for Incomplete Decision System

**Tao Yan and Chongzhao Han**

*Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University,  
No. 28 Xianning West Road, Xi'an, Shaanxi 710049, China*

Correspondence should be addressed to Tao Yan; [tao.yan@stu.xjtu.edu.cn](mailto:tao.yan@stu.xjtu.edu.cn)

Received 16 September 2013; Accepted 19 December 2013; Published 12 February 2014

Academic Editor: Panos Liatsis

Copyright © 2014 T. Yan and C. Han. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pawlak's classical rough set theory has been applied in analyzing ordinary information systems and decision systems. However, few studies have been carried out on the attribute selection problem in incomplete decision systems because of its complexity. It is therefore necessary to investigate effective algorithms to deal with this issue. In this paper, a new rough conditional entropy-based uncertainty measure is introduced to evaluate the significance of subsets of attributes in incomplete decision systems. Furthermore, some important properties of rough conditional entropy are derived and three attribute selection approaches are constructed, including an exhaustive search strategy approach, a heuristic search strategy approach, and a probabilistic search strategy approach for incomplete decision systems. Moreover, several experiments on real-life incomplete data sets are conducted to assess the efficiency of the proposed approaches. The final experimental results indicate that two of these approaches can give satisfying performances in the process of attribute selection in incomplete decision systems.

## 1. Introduction

Rough set theory, proposed by Pawlak [1–3], is an extension of set theory for the study of the intelligent systems characterized by uncertain, imprecise, incomplete, and inconsistent data. It has been proven to be an innovative and efficient mathematic tool, compared with other traditional data processing strategies like PCA, neural networks, SVM and so forth [4–7]. Unlike those methods, rough set theory allows knowledge discovering process to be conducted automatically by the data themselves without any dependence on the prior knowledge. By using the concepts of upper and lower approximations in rough set model to deal with the data, knowledge hidden in information systems could be discovered and expressed in the form of decision rules. Rough set methodology presents a novel paradigm to deal with uncertainty and then it has been successfully applied into feature selection [8], rule extraction [9, 10], uncertainty reasoning [11–13], decision evaluation [14], granular computing [15, 16], and so on.

It has been known that Pawlak's classical rough set model can only be used to tackle the problems of complete

information systems [17]. Nevertheless, because of the error of data measuring, the impreciseness from the limitation of data acquisition manners and other probable factors, in the real database it is inevitable to meet the empty values, which stand for the inaccessible information in the database for the moment. In other words, the incomplete information systems with missing values often exist in practical knowledge acquisition. For now, two main approaches have been proposed to cope with incomplete information systems. One is indirect approach, which transforms an incomplete information system into a complete information system by eliminating objects with missing values or filling up missing values with processed data. The other is direct approach, which extends some basic notions in the classical rough set models [18, 19]. In the past decade, with respect to different requirements, various extensions of the rough set models have been proposed, such as variable precision rough set models [20, 21], rough set models based on tolerance relation [22, 23], fuzzy rough set models [24], and weighted attributes values rough set models [25].

In many fields such as data mining, machine learning, and pattern recognition, data sets containing huge numbers

of attributes can often be encountered. In such cases, feature selection, or attribute reduction as we know, is necessary. It is well known that irrelevant and redundant attributes in input attributes not only complicate the problem but also degrade solution accuracy [26, 27]. As a significant step among data preprocessing procedures, the main objective of attribute selection is to determine a minimal attribute subset, which is also called reduction, from a problem domain, which can retain a relatively high accuracy in replacing the original attributes.

In comparison with the study of attribute selection from complete information systems, scant effort has been made to develop the tolerance relation-based methods of attribute selection for incomplete information systems [28–31]. Furthermore, there is a lack of the study of attribute selection of incomplete decision systems, compared with the above. As an effective method for attribute selection, rough set can preserve the meaning of the attributes. The essence of rough set approach for attribute selection is to find a subset of attributes, which can predict the decision concepts as well as the original attribute set. So far, there have been some new widespread approaches which are usually considered to develop the classical rough set theory as follows: tolerance-based rough set model [32], covering rough set model [33, 34], and dominance-based rough set model [35, 36]. However, they still have some inherent disadvantages and are not suitable for attribute selection of incomplete decision systems.

As we know, finding the minimal reduct in an incomplete decision system is an NP-hard problem [37]. The general method of solving this kind of problems is adopting a heuristic search, which always depends on the measurements associated with the attributes [38]. However, little investigation has addressed the issue of measuring the uncertainty of knowledge of the tolerance relation-based rough set models in incomplete decision system until now. Hence, a further study on uncertainty measures applicable for evaluating the roughness and accuracy of a set in an incomplete decision system is of both theoretical and practical importance.

The main aim of this paper is to construct an effective uncertainty measure evaluating the roughness and accuracy of knowledge to find a heuristic attribute selection algorithm of incomplete decision systems. The rest of this paper is organized as follows. In Section 2, we briefly review some fundamental concepts concerning the main subject of this paper. Section 3 introduces the concept of entropy-based uncertainty measures and demonstrates several heuristic attribute selection algorithms of incomplete decision systems. Experimental comparisons and results are illustrated and analyzed in Section 4. Finally, some conclusions are presented in Section 5.

## 2. Preliminary

Classical rough set theory is originated by Qian et al. to deal with imprecise or vague concepts [39]. In the last decade, many generalized rough set models have been proposed

and developed. In this section, we will only introduce some notions being used in this paper.

Pawlak's classical rough set is to be defined at the first place. An information system in rough set theory is a pair  $(U, A)$ , where  $U = \{x_1, \dots, x_n\}$  denotes a nonempty finite set of objects, called universe of discourse, and  $A = \{a_1, \dots, a_m\}$  denotes a finite condition attribute set. With every attribute  $a \in A$  we associate a set  $V_a$ , of its values, called the domain of  $a$ . Then, a data pattern  $(a_1(x), \dots, a_m(x))$  can be defined by the object  $x$  and attributes from  $A$ . A decision table, abbreviated as DT, is a special system with the form  $(U, C \cup \{d\})$ , where  $d \notin C$  denotes decision attribute. Let  $V_d$  denote the domain of decision attribute mapping  $d(x)$ . For the application of pattern classification, the attribute set is just the feature set and the universe of the discourse may represent a training pattern set or a sign set of training pattern sets.

Let  $R$  be an equivalence relation on  $U$ , which means relation  $R$  satisfies reflexivity, symmetry, and transitivity. Relation  $R$  generates a partition  $U/R = \text{IND}(R) = \{\{x\}_R \mid x \in U\}$  on  $U$ , where  $\text{IND}(R)$  denotes the equivalence classes, as well as indiscernible class, generated by the equivalence relation  $R$ . These are also called elementary sets of  $R$  in rough set theory. Let  $\emptyset$  denote the empty sets. For any  $X \subseteq U$ , we can describe  $X$  by elementary sets of  $R$  and the two sets

$$R_-(X) = \{\{x\}_R \mid \{x\}_R \subseteq X\}, \quad (1)$$

$$R^-(X) = \{\{x\}_R \mid \{x\}_R \cap X \neq \emptyset\}$$

which are called the lower and upper approximations of  $X$ , respectively. Then, the concepts of positive region,

$$\text{POS}(X) = R_-(X), \quad (2)$$

negative region,

$$\text{NEG}(X) = U - R^-(X), \quad (3)$$

boundary region,

$$\begin{aligned} \text{BN}(X) &= R^-(X) - R_-(X) \\ &= \{\{x\}_R \mid \{x\}_R \cap X \neq \emptyset\} - \{\{x\}_R \mid \{x\}_R \subseteq X\}, \end{aligned} \quad (4)$$

and approximation measure,

$$\alpha_R(X) = \frac{|R_-(X)|}{|R^-(X)|}, \quad (5)$$

are introduced, where  $X \neq \emptyset$ . The lower approximation, which is equivalent to the positive region, is the complete set of the objects in the universe that can be unambiguously classified as belonging to the target set  $X$ . In contrast, the upper approximation is the complete set of the objects that are possibly members of the target set  $X$ . In other words, these objects cannot be positively classified as belonging to the complement of the target set  $X$ , that is,  $\bar{X}$ . Furthermore, the negative region contains the set of objects that can be definitely ruled out as the members of the target set  $X$ . Finally, the approximation measure  $\alpha_R(X)$  is intended to capture

the degree of completeness of our knowledge about the target set  $X$ .

In most cases, some precise values of particular attributes in an information system are not known, which means missing or known partially. Then such a system is called an incomplete information system and is still denoted without any confusion by a pair  $(U, A)$ . As for an incomplete information system, a missing value  $a(x)$  may be represented by the set of all the possible values for the corresponding attribute; that is,  $a(x) = V_a$ . Moreover, if  $a(x)$  is known partially, for instance, which implies that  $a(x)$  cannot be  $b, c \in V_a$ , then the value of  $a(x)$  is specified as  $V_a - \{b, c\}$ .

In what follows, we only take the consideration of the incomplete information systems with missing values. In such a case, the special symbol “\*” is to be used to indicate that the specific value of an attribute is missing. Let  $S = (U, AT)$  be an incomplete information system. Each subset of attributes  $A \subseteq AT$  determines a similarity relation

$$\begin{aligned} \text{SIM}(A) &= \{(x, y) \in U \times U \mid \forall a \in A, \\ & a(x) = a(y) \cup a(x) = * \cup a(y) = *\}. \end{aligned} \quad (6)$$

The incomplete information system can be described by a set-valued information system with  $a(x) = V_a$  when  $a(x) = *$ . In such a case, the similarity relation  $\text{SIM}(A)$  can be equivalently defined as

$$\text{SIM}(A) = \{(x, y) \in U \times U \mid \forall a \in A, a(x) \cap a(y) \neq \emptyset\}. \quad (7)$$

With the similarity relation  $\text{SIM}(A)$ , two objects are considered to be possibly indiscernible in terms of the values of attributes  $A$ . Furthermore, a similarity relation satisfies reflexive and symmetric, but may not be transitive, so it is one of the tolerance relations.

### 3. Rough Entropy-Based Uncertainty Measures

In this section, the concept of rough entropy is introduced to measure the uncertainty or roughness of knowledge in incomplete information systems. And then, some rough entropy-based uncertainty measures are presented in incomplete information systems and incomplete decision systems. Some important properties concerning the uncertainty measures are derived, respectively, and the relationship among them is discussed as well.

For a given information system, we need to assess its uncertainty or roughness for a target object or a target decision. A form of uncertainty measure, which is called rough entropy, has been mentioned in rough sets, rough relational databases, and information systems to calculate the roughness of knowledge. The following definition gives the description of the rough entropy in incomplete information systems.

**3.1. Rough Entropy in IIS and IDS.** Let  $\text{IIS} = (U, A)$  denote an incomplete information system with  $* \in V$  and  $P \subseteq A$ , and  $U/\text{SIM}(P) = \{S_P(u_1), S_P(u_2), \dots, S_P(u_{|U|})\}$ . According to

Shannon's theory of information entropy, the rough entropy of knowledge  $P$  on  $U$  is denoted by

$$\text{EN}(P) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{1}{|S_P(u_i)|}, \quad (8)$$

where  $1/|S_P(u_i)|$  represents the probability of an element within the tolerance class  $S_P(u_i)$  [40].

Let  $\text{IIS} = (U, A)$  be an incomplete information system, with  $P, Q \subseteq A$ . If there exists a one-to-one onto function  $h : U/\text{SIM}(P) \rightarrow U/\text{SIM}(Q)$  such that  $|h(S_P(u_i))| = |S_Q(u_i)| = |S_P(u_i)|$  for any  $u_i \in U$ , then  $\text{EN}(P) = \text{EN}(Q)$ . Therefore, the rough entropy of knowledge is invariant with respect to the different set of tolerance classes that are size isomorphic.

Let  $\text{IDS} = (U, C \cup \{d\})$  denote an incomplete decision system with  $* \in V_C$ . For any subset of condition attributes  $B \subseteq C$ , a tolerance relation  $T$ , which is in generalized form of the former similarity relation, can be defined as

$$\begin{aligned} T_B &= \{(x, y) \in U \times U \mid \forall a \in B, \\ & a(x) = a(y) \cup a(x) = * \cup a(y) = *\}. \end{aligned} \quad (9)$$

Then, the tolerance class of an object  $x$  with respect to an attribute set  $B$  is defined as

$$T_B(x) = \{y \in U \mid (x, y) \in T_B\}. \quad (10)$$

Obviously, relation  $T$  is reflexive and symmetric, but may not be transitive.

**Definition 1.** Let  $\text{IDS} = (U, C \cup \{d\})$ ,  $* \in V_C$  be an incomplete decision system, then the generalized decision  $\partial_B = U \rightarrow V_d$  is defined as follows:

$$\partial_B(x) = \{u \mid y \in T_B(x) \cap u = d(y)\}. \quad (11)$$

**Definition 2.** Let  $\text{IDS} = (U, C \cup \{d\})$ ,  $* \in V_C$  be an incomplete decision system. Let  $\partial_B = U \rightarrow V_d$  be the generalized decision function. If  $|\partial_B(x)| = 1$  for any  $x \in U$ , then the incomplete decision is consistent, which implies that it is deterministic and definite, where  $|\cdot|$  denotes the number of the elements of the set.

In what follows, we give an example, which was firstly shown by Kryszkiewicz [31], to demonstrate how tolerance relationship works in a specific incomplete decision system. Given an incomplete decision system shown in Table 1, where  $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$ ,  $C = \{\text{Price, Mileage, Size, Max-speed}\} = \{P, M, S, X\}$

TABLE 1: An example of incomplete decision system.

Car	Price	Mileage	Size	Max-speed	Decision
1	High	High	Full	Low	Good
2	Low	*	Full	Low	Good
3	*	*	Compact	High	Poor
4	High	*	Full	High	Good
5	*	*	Full	High	Excellent
6	Low	High	Full	*	Good

and  $d = \text{Acceleration}$ , and  $U/\text{IND}(d) = \{S_{\text{Good}}, S_{\text{Poor}}, S_{\text{Excel}}\}$ . By definition, we have

$$\begin{aligned}
T_C(u_1) &= \{u_1\}, & T_C(u_1 \cap d_{\text{Good}}) &= \{u_1\}, \\
T_C(u_1 \cap d_{\text{Poor}}) &= T_C(u_1 \cap d_{\text{Excellent}}) = \emptyset, \\
T_C(u_2) &= \{u_2, u_6\}, & T_C(u_2 \cap d_{\text{Good}}) &= \{u_2, u_6\}, \\
T_C(u_2 \cap d_{\text{Poor}}) &= T_C(u_2 \cap d_{\text{Excellent}}) = \emptyset, \\
T_C(u_3) &= \{u_3\}, & T_C(u_3 \cap d_{\text{Poor}}) &= \{u_3\}, \\
T_C(u_3 \cap d_{\text{Good}}) &= T_C(u_3 \cap d_{\text{Excellent}}) = \emptyset, \\
T_C(u_4) &= \{u_4, u_5\}, & T_C(u_4 \cap d_{\text{Good}}) &= \{u_4\}, \\
T_C(u_4 \cap d_{\text{Poor}}) &= \emptyset, & T_C(u_4 \cap d_{\text{Excellent}}) &= \{u_5\}, \\
T_C(u_5) &= \{u_4, u_5, u_6\}, & T_C(u_5 \cap d_{\text{Good}}) &= \{u_4, u_6\}, \\
T_C(u_5 \cap d_{\text{Poor}}) &= \emptyset, & T_C(u_5 \cap d_{\text{Excellent}}) &= \{u_5\}, \\
T_C(u_6) &= \{u_2, u_5, u_6\}, & T_C(u_6 \cap d_{\text{Good}}) &= \{u_2, u_6\}, \\
T_C(u_6 \cap d_{\text{Poor}}) &= \emptyset, & T_C(u_6 \cap d_{\text{Excellent}}) &= \{u_5\}, \\
\frac{U}{\text{IND}(d)} &= \{\{u_1, u_2, u_4, u_6\}, \{u_3\}, \{u_5\}\}.
\end{aligned} \tag{12}$$

Consequently, we get  $|\partial_C(u_1)| = |\partial_C(u_2)| = |\partial_C(u_3)| = |\partial_C(u_4)| = |\partial_C(u_5)| = |\partial_C(u_6)| = 1$ , which proves that the incomplete decision system is consistent.

**Definition 3.** Let  $\text{IDS} = (U, C \cup \{d\})$  be an incomplete decision system. Let  $B \subseteq C$ , then the attribute set  $B$  is a relative reduct of  $\text{IDS}$ , if and only if

- (1)  $\partial_B(x) = \partial_A(x)$  for all  $x \in U$ ,
- (2)  $b \in B, \partial_{B-\{b\}} \neq \partial_C$

are satisfied simultaneously.

**3.2. Conditional Entropy Measure for Incomplete Decision System.** The rough entropy of knowledge in IIS and IDS has been discussed above. In this subsection, we will introduce a new form of conditional entropy and the mutual information based on the tolerance relationship to measure the uncertainty of knowledge in incomplete decision systems. And then, some important properties will be deduced.

**Definition 4.** Given a consistent incomplete decision system,  $\text{IDS} = (U, C \cup \{d\})$  and  $B \subseteq C$ . Let  $U/T_B = \{T_B(x_1), T_B(x_2), \dots, T_B(x_{|U|})\}$ ,  $U/d = \{d_1, d_2, \dots, d_m\}$ . The conditional entropy of  $B$  to  $d$  is defined as follows:

$$\begin{aligned}
\text{EN}(d | B) &= -\sum_{i=1}^{|U|} p(T_B(u_i)) \\
&\quad \times \sum_{j=1}^{|U/d|} p(d_j | T_B(u_i)) \log_2 p(d_j | T_B(u_i)),
\end{aligned} \tag{13}$$

where

$$\begin{aligned}
p(T_B(u_i)) &= \frac{|T_B(u_i)|}{|U|}, \quad i = 1, 2, \dots, |U|, \\
p(d_j | T_B(u_i)) &= \frac{p(T_B(u_i), d_j)}{p(T_B(u_i))} = \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|}, \\
&\quad i = 1, 2, \dots, |U|, \quad j = 1, 2, \dots, m.
\end{aligned} \tag{14}$$

Hence, we have

$$\begin{aligned}
\text{EN}(d | B) &= -\sum_{i=1}^{|U|} p(T_B(x_i)) \sum_{j=1}^{|U/d|} p(d_j | T_B(x_i)) \log_2 p(d_j | T_B(x_i)) \\
&= -\sum_{i=1}^{|U|} \frac{|T_B(x_i)|}{|U|} \sum_{j=1}^{|U/d|} \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \\
&= \sum_{i=1}^{|U|} \sum_{j=1}^{|U/d|} \frac{|T_B(x_i) \cap d_j|}{|U|} \log_2 \frac{|T_B(x_i)|}{|T_B(x_i) \cap d_j|}.
\end{aligned} \tag{15}$$

It is obvious that  $\text{EN}(d | B) = 0$  when  $T_B(x_i) \cap d_j = \emptyset$ .

**Proposition 5.** Let  $\text{IDS} = (U, C \cup \{d\})$  be a consistent incomplete decision system. Then, we have  $\text{EN}(d | C) = 0$ .

*Proof.* Since  $\text{IDS}$  is consistent, we have  $\partial_C(x_i) = 1, x_i \in U$ . This means that  $T_C(x_i) \subseteq d_j, d_j \in U/d, x_i \in U$ .

Hence, we have

$$T_C(x_i) \cap d_r = \emptyset, \quad \forall d_r \in \frac{U}{d}, d_r \neq d_j. \tag{16}$$

Consequently, we know that

$$\begin{aligned}
& EN(d | C) \\
&= -\sum_{i=1}^{|U|} p(T_C(x_i)) \sum_{j=1}^{|U/d|} p(d_j | T_C(x_i)) \log_2 p(d_j | T_C(x_i)) \\
&= -\sum_{i=1}^{|U|} \frac{|T_C(x_i)|}{|U|} \sum_{j=1}^{|U/d|} \frac{|T_C(x_i) \cap d_j|}{|T_C(x_i)|} \log_2 \frac{|T_C(x_i) \cap d_j|}{|T_C(x_i)|} \\
&= -\sum_{i=1}^{|U|} \frac{|T_C(x_i)|}{|U|} (0 + 0 + \dots + 1 \log 1 + 0 + \dots + 0) = 0.
\end{aligned} \tag{17}$$

**Proposition 6.** Let  $IDS = (U, C \cup \{d\})$  be a consistent incomplete decision system;  $B \subseteq C$  is a relative reduct of  $C$  relative to decision attribute  $d$ , if and only if

- (1)  $EN(d | C) = EN(d | B)$ ;
- (2)  $B' \subset B$ ,  $EN(d | B') \neq EN(d | B)$ .

*Proof.* (1) We first prove that if  $EN(d | C) = EN(d | B)$ ,  $\forall B' \subset B$ ,  $EN(d | B') \neq EN(d | B)$  then  $B \subseteq C$  is a relative reduct.

By Proposition 5, we know that  $EN(d | C) = 0$ . Then  $EN(d | B) = EN(d | C) = 0$ .

Since

$$0 \leq \frac{|T_B(x_i)|}{|U|} \leq 1, \quad 0 \leq \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \leq 1, \tag{18}$$

we have

$$\frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \leq 0. \tag{19}$$

At the same time, we know that  $EN(d | B) = 0$  if and only if

$$\frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} = 0, \quad \forall x_i, d_j. \tag{20}$$

This implies that there exists only one  $d_j \in U/d$  such that  $T_B(x_i) \subseteq d_j$  and  $T_B(x_i) \cap d_r = \emptyset$ ,  $\forall d_r \in U/d$ ,  $d_r \neq d_j$ . This means that  $|\partial_B(x_i)| = 1$ ,  $\forall x_i \in U$ . Hence, we get  $\partial_B = \partial_C$ .

From  $\forall B' \subset B$ ,  $EN(d | B') \neq EN(d | B)$ , we have  $EN(d | B') \neq 0$ . It follows that there at least exist  $x_i$ ,  $d_j$ , and  $d_k$  such that  $T_{B'}(x_i) \cap d_j \neq \emptyset$  and  $T_{B'}(x_i) \cap d_k \neq \emptyset$ . This means that  $|\partial_{B'}(x_i)| \geq 2$ . Hence,  $\partial_{B'} \neq \partial_B$ .

(2) We then prove that if  $B \subseteq C$  is a relative reduct, then  $EN(d | C) = EN(d | B)$ ,  $\forall B' \subset B$ ,  $EN(d | B') \neq EN(d | B)$  hold.

Since  $B \subseteq C$  is a relative reduct, we get  $\partial_B = \partial_C$ ; that is,  $|\partial_B(x_i)| = |\partial_C(x_i)| = 1$ ,  $\forall x_i \in U$ . It follows that,  $\forall x_i \in U$ , there exists  $d_j \in U/d$  such that  $T_B(x_i) \subseteq d_j$  and  $T_B(x_i) \cap d_r = \emptyset$ ,  $\forall d_r \in U/d$ ,  $d_r \neq d_j$ .

It follows that

$$\begin{aligned}
\frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} &= \frac{|T_B(x_i)|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i)|}{|T_B(x_i)|} \\
&= 1 \log 1 = 0, \quad \forall x_i, d_j.
\end{aligned} \tag{21}$$

Therefore, this means that

$$\sum_{j=1}^{|U/d|} \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} = 0, \quad \forall x_i. \tag{22}$$

Consequently, we know

$$\begin{aligned}
& EN(d | B) \\
&= -\sum_{i=1}^{|U|} p(T_B(x_i)) \sum_{j=1}^{|U/d|} p(d_j | T_B(x_i)) \log_2 p(d_j | T_B(x_i)) \\
&= -\sum_{i=1}^{|U|} \frac{|T_B(x_i)|}{|U|} \sum_{j=1}^{|U/d|} \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \log_2 \frac{|T_B(x_i) \cap d_j|}{|T_B(x_i)|} \\
&= -\sum_{i=1}^{|U|} \frac{|T_B(x_i)|}{|U|} \cdot 0 = 0.
\end{aligned} \tag{23}$$

For  $\forall B' \subset B$ , we know that  $\partial_{B'} \neq \partial_B$ ; that is, there at least exists  $x_i \in U$  such that  $\partial_{B'}(x_i) \neq \partial_B(x_i)$ . It follows that  $|\partial_{B'}(x_i)| > |\partial_B(x_i)| = 1$ . Thus, there at least exist  $d_j \in U/d$  and  $d_r \in U/d$ ,  $d_r \neq d_j$  such that

$$\begin{aligned}
& T_{B'}(x_i) \cap d_j \neq \emptyset, \\
& T_{B'}(x_i) \cap d_r \neq \emptyset.
\end{aligned} \tag{24}$$

Then, we get  $EN(d | B') \neq EN(d | B)$ .

From parts (1) and (2), we finally prove that the theorem holds.  $\square$

#### 4. Attribute Selection Approaches Based on Rough Conditional Entropy for IDS

Two important steps are contained in the procedure of attribute selection: evaluation of a candidate attribute subset and search strategy through the attribute space, which are all for finding the most significant attributes, that is, relative reduct. Therefore, we use the conditional entropy discussed previously in Section 3 to evaluate the attribute subset and a measurement is defined as follows.

*Definition 7.* Given a consistent decision system  $IDS = (U, C \cup \{d\})$ , let  $B \subseteq C$  and  $a \in C - B$ . Then, the significance of attribute relative to  $B$  is defined as

$$SIG(a, B, d) = EN(d | B) - EN(d | B \cup \{a\}). \tag{25}$$

This definition describes the increment of discernibility power relative to the decision caused by involving attribute  $a$ .



It implies that the larger the difference between  $EN(d | B)$  and  $EN(d | B \cup \{a\})$  is, the more significant the specific condition attribute  $a$  is for condition attribute subset  $B$ . Thus, it can be used as a new measurement for attribute selection in incomplete decision system. According to this new measurement, three attribute selection approaches based on different search strategies are proposed, respectively, in the following subsections.

**4.1. Breadth-First: Exhaustive Search.** Breadth-first is one of earliest feature selection or attribute selection algorithms in machine learning area. It begins with an empty attribute set and carries out search process with breadth-first strategy until it finds a minimal subset that satisfies stop criterion. Since the Breadth-first algorithm adopts an exhaustive search strategy, it can guarantee an optimal solution [41]. Then, we present the Breadth-first approach for attribute selection of incomplete decision system as shown in Algorithm 1.

*Example 8.* Given the incomplete decision system shown in Table 2, we have

$$C = \{P, M, S, X\},$$

$$\frac{U}{d} = \{\{u_1, u_2, u_4, u_6\}, \{u_3\}, \{u_5\}\},$$

$$T_C(u_1) = T_C(u_2) = T_C(u_6) = \{u_1, u_2, u_6\}, \quad (26)$$

$$T_C(u_3) = \{u_3\},$$

$$T_C(u_4) = \{u_4\},$$

$$T_C(u_5) = \{u_5\}.$$

According to the previous definition of rough conditional entropy, we can simply get

$$\begin{aligned} EN(d | C) = & - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + \frac{1}{6} * (1 \log_2 1) \right. \\ & \left. + \frac{1}{6} * (1 \log_2 1) + \frac{1}{6} * (1 \log_2 1) \right] = 0. \end{aligned} \quad (27)$$

Consequently, we can deduce other conditional entropy values with respect to the different condition attributes in the same way as follows:

$$T_P(u_1) = T_P(u_4) = \{u_1, u_2, u_3, u_4, u_6\},$$

$$T_P(u_2) = T_P(u_3) = T_P(u_6) = \{u_1, u_2, u_3, u_4, u_5, u_6\},$$

$$T_P(u_5) = \{u_2, u_3, u_5, u_6\},$$

$$EN(d | \{P\})$$

$$\begin{aligned} = & - \left[ 2 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) \right. \\ & + 3 * \frac{6}{6} * \left( \frac{4}{6} \log_2 \frac{4}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6} \right) \\ & \left. + \frac{4}{6} * \left( \frac{2}{4} \log_2 \frac{2}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \right] \\ = & 5.9581, \end{aligned}$$

$$T_M(u_1) = T_M(u_2) = T_M(u_3) = T_M(u_4)$$

$$= T_M(u_5) = T_M(u_6) = \{u_1, u_2, u_3, u_4, u_5, u_6\},$$

$$EN(d | \{M\})$$

$$\begin{aligned} = & - \left[ 6 * \frac{6}{6} * \left( \frac{4}{6} \log_2 \frac{4}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6} \right) \right] \\ = & 7.5098, \end{aligned}$$

$$T_S(u_1) = T_S(u_2) = T_S(u_4) = T_S(u_5)$$

$$= T_S(u_6) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$T_S(u_3) = \{u_3\},$$

$$EN(d | \{S\})$$

$$\begin{aligned} = & - \left[ 5 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) + \frac{1}{6} * (1 \log_2 1) \right] \\ = & 3.0080, \end{aligned}$$

$$T_X(u_1) = T_X(u_2) = T_X(u_6) = \{u_1, u_2, u_6\},$$

$$T_X(u_3) = T_X(u_4) = T_X(u_5) = \{u_3, u_4, u_5\},$$

$$EN(d | \{X\})$$

$$\begin{aligned} = & - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + 3 * \frac{3}{6} \right. \\ & \left. * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] \\ = & 2.3774. \end{aligned} \quad (28)$$

Then, we calculate the rough conditional entropy values with respect to different combinations of two condition attributes:

$$T_{P,M}(u_1) = T_{P,M}(u_4) = \{u_1, u_2, u_3, u_4, u_6\},$$

$$T_{P,M}(u_2) = T_{P,M}(u_3) = T_{P,M}(u_6) = \{u_1, u_2, u_3, u_4, u_5, u_6\},$$

$$T_{P,M}(u_5) = \{u_2, u_3, u_5, u_6\},$$

$$EN(d | \{P, M\})$$

$$= - \left[ 2 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) \right]$$

Input: An incomplete decision system  $(U, C \cup \{d\})$ .  
Output: An attribute selection result  $R$ .

- (1) For every size = 0 to  $|C|$
- (2) For all subset SelectAttr with  $|\text{SelectAttr}| = \text{size}$
- (3) If  $\text{EN}(d | \text{SelectAttr}) \neq \text{EN}(d | C)$ , go to Step 2, otherwise return  $R = \text{SelectAttr}$
- (4) End
- (5) End

ALGORITHM 1: Breadth-first.

TABLE 2: Another example of incomplete decision system.

Car	Price	Mileage	Size	Max-speed	Decision
1	High	High	Full	Low	Good
2	*	*	Full	Low	Good
3	*	*	Compact	High	Poor
4	High	*	Full	High	Good
5	Low	*	Full	High	Excellent
6	*	High	Full	Low	Good

$$\begin{aligned}
& + 3 * \frac{6}{6} * \left( \frac{4}{6} \log_2 \frac{4}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6} \right) \\
& + \frac{4}{6} * \left( \frac{2}{4} \log_2 \frac{2}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \\
& = 5.9581,
\end{aligned}$$

$$T_{P,S}(u_1) = T_{P,S}(u_4) = \{u_1, u_2, u_4, u_6\},$$

$$T_{P,S}(u_2) = T_{P,S}(u_6) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$T_{P,S}(u_3) = \{u_3\},$$

$$T_{P,S}(u_5) = \{u_2, u_5, u_6\},$$

$\text{EN}(d | \{P, S\})$

$$\begin{aligned}
& = - \left[ 2 * \frac{4}{6} * \left( \frac{4}{4} \log_2 \frac{4}{4} \right) \right. \\
& \quad + 2 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) \\
& \quad \left. + \frac{1}{6} * (1 \log_2 1) + \frac{3}{6} * \left( \frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] \\
& = 1.6624,
\end{aligned}$$

$$T_{P,X}(u_1) = T_{P,X}(u_2) = T_{P,X}(u_6) = \{u_1, u_2, u_6\},$$

$$T_{P,X}(u_3) = \{u_3, u_4, u_5\},$$

$$T_{P,X}(u_4) = \{u_3, u_4\},$$

$$T_{P,X}(u_5) = \{u_3, u_5\},$$

$\text{EN}(d | \{P, X\})$

$$= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) \right]$$

$$+ \frac{3}{6} * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right)$$

$$+ \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right)$$

$$+ \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) = 1.4591,$$

$$T_{M,S}(u_1) = T_{M,S}(u_2) = T_{M,S}(u_4) = T_{M,S}(u_5)$$

$$= T_{M,S}(u_6) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$T_{M,S}(u_3) = \{u_3\},$$

$\text{EN}(d | \{M, S\})$

$$\begin{aligned}
& = - \left[ 5 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) + \frac{1}{6} * (1 \log_2 1) \right] \\
& = 3.0080,
\end{aligned}$$

$$T_{M,X}(u_1) = T_{M,X}(u_2) = T_{M,X}(u_6) = \{u_1, u_2, u_6\},$$

$$T_{M,X}(u_3) = T_{M,X}(u_4) = T_{M,X}(u_5) = \{u_3, u_4, u_5\},$$

$\text{EN}(d | \{M, X\})$

$$\begin{aligned}
& = - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + 3 * \frac{3}{6} \right. \\
& \quad \left. * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] \\
& = 2.3774,
\end{aligned}$$

$$T_{S,X}(u_1) = T_{S,X}(u_2) = T_{S,X}(u_6) = \{u_1, u_2, u_6\},$$

$$T_{S,X}(u_3) = \{u_3\},$$

$$T_{S,X}(u_4) = T_{S,X}(u_5) = \{u_4, u_5\},$$

$$\begin{aligned}
& \text{EN}(d \mid \{S, X\}) \\
&= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + \frac{1}{6} * (1 \log_2 1) \right. \\
&\quad \left. + 2 * \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \right] = 0.6667.
\end{aligned} \tag{29}$$

Finally, we calculate the rough conditional entropy values with respect to different combinations of three condition attributes:

$$T_{P,M,S}(u_1) = T_{P,M,S}(u_4) = \{u_1, u_2, u_4, u_6\},$$

$$T_{P,M,S}(u_2) = T_{P,M,S}(u_6) = \{u_1, u_2, u_4, u_5, u_6\},$$

$$T_{P,M,S}(u_3) = \{u_3\},$$

$$T_{P,M,S}(u_5) = \{u_2, u_5, u_6\},$$

$$\begin{aligned}
& \text{EN}(d \mid \{P, M, S\}) \\
&= - \left[ 2 * \frac{4}{6} * \left( \frac{4}{4} \log_2 \frac{4}{4} \right) \right. \\
&\quad + 2 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) + \frac{1}{6} * (1 \log_2 1) \\
&\quad \left. + \frac{3}{6} * \left( \frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] = 1.6624,
\end{aligned}$$

$$T_{P,M,X}(u_1) = T_{P,M,X}(u_2) = T_{P,M,X}(u_6) = \{u_1, u_2, u_6\},$$

$$T_{P,M,X}(u_3) = \{u_3, u_4, u_5\},$$

$$T_{P,M,X}(u_4) = \{u_3, u_4\},$$

$$T_{P,M,X}(u_5) = \{u_3, u_5\},$$

$$\begin{aligned}
& \text{EN}(d \mid \{P, M, X\}) \\
&= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) \right. \\
&\quad + \frac{3}{6} * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \\
&\quad + \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \\
&\quad \left. + \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \right] = 1.4591,
\end{aligned}$$

$$T_{P,S,X}(u_1) = T_{P,S,X}(u_2) = T_{P,S,X}(u_6) = \{u_1, u_2, u_6\},$$

$$T_{P,S,X}(u_3) = \{u_3\},$$

$$T_{P,S,X}(u_4) = \{u_4\},$$

$$T_{P,S,X}(u_5) = \{u_5\},$$

$$\begin{aligned}
& \text{EN}(d \mid \{P, S, X\}) \\
&= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + \frac{1}{6} * (1 \log_2 1) \right. \\
&\quad \left. + \frac{1}{6} * (1 \log_2 1) + \frac{1}{6} * (1 \log_2 1) \right] = 0.
\end{aligned} \tag{30}$$

Since  $\text{EN}(d \mid \{P, S, X\}) = \text{EN}(d \mid C) = 0$ , we obtain the desired attribute set of selected  $\{P, S, X\}$ , which is a relative reduct of the original condition attribute set  $C$ . It also means that the search procedure ends at this step and it is not necessary to calculate the entropy value of the last combination of three attributes  $\{M, S, X\}$  anymore. The detailed search procedure is illustrated in Figure 1.

**4.2. Depth-First: Heuristic Search.** We can also use heuristic search or greed search to find attribute reduction. At the very beginning, the candidate attribute subset is empty. Then, a new attribute which can maximize the significance measure is added to the selected attribute subset each time, until the stop criterion is satisfied. Depth-first algorithm is fast, close to optimal, and deterministic [42]. Here, we present the Depth-first approach for incomplete decision system, as shown in Algorithm 2.

*Example 9.* Given the incomplete decision system shown in Table 2, we have

$$\text{SelectAttr} = \emptyset,$$

$$\text{unSelectAttr} = C = \{P, M, S, X\},$$

$$\frac{U}{d} = \{\{u_1, u_2, u_4, u_6\}, \{u_3\}, \{u_5\}\},$$

$$\begin{aligned}
& \text{EN}(d \mid \{P\}) \\
&= - \left[ 2 * \frac{5}{6} * \left( \frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right) \right. \\
&\quad + 3 * \frac{6}{6} * \left( \frac{4}{6} \log_2 \frac{4}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6} \right) \\
&\quad \left. + \frac{4}{6} * \left( \frac{2}{4} \log_2 \frac{2}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \right] \\
&= 5.9581,
\end{aligned}$$

$$\begin{aligned}
& \text{EN}(d \mid \{M\}) \\
&= - \left[ 6 * \frac{6}{6} * \left( \frac{4}{6} \log_2 \frac{4}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6} \right) \right] \\
&= 7.5098,
\end{aligned}$$

$$\begin{aligned}
& \text{EN}(d \mid \{S\}) \\
&= - \left[ 5 * \frac{5}{6} * \left( \frac{4}{5} * \log_2 \frac{4}{5} + \frac{1}{5} * \log_2 \frac{1}{5} \right) + \frac{1}{6} * (1 \log_2 1) \right] \\
&= 3.0080,
\end{aligned}$$



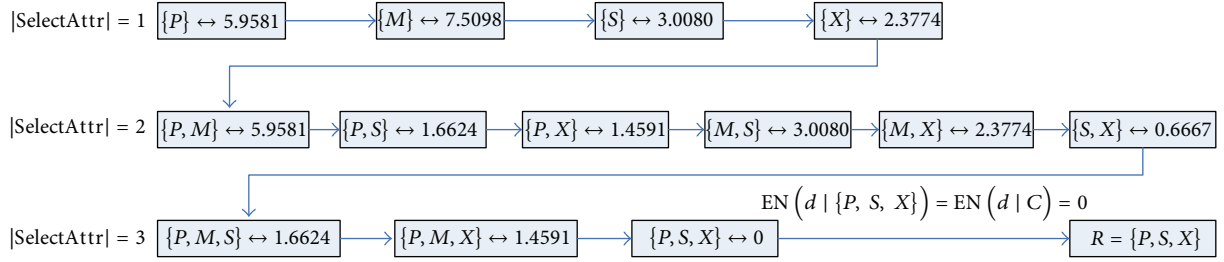


FIGURE 1: The flow chart of the Breadth-first algorithm.

Input: An incomplete decision system  $(U, C \cup \{d\})$ .

Output: A selected attribute subset  $R$ .

- (1) Initialize  $\text{SelectAttr} = \emptyset$ ,  $\text{unSelectAttr} = C$
- (2) For every attribute  $a_i \in \text{unSelectAttr}$ ,  $\text{SelectAttr} = \text{SelectAttr} \cup \{a_i\}$ ,  
calculate the tolerance class of object  $u_i$ ,  $T_{\text{Attr}}(u_i)$
- (3) Calculate the conditional entropy  $\text{EN}(d | \text{SelectAttr}) = \text{EN}(d | \text{SelectAttr} \cup \{a_i\})$
- (4) Choose the attribute  $a$  which minimizes  $\text{EN}(d | \text{SelectAttr} \cup \{a\})$ , that is  
choose the attribute with the maximum significance measure  $\text{SIG}(a, \text{SelectAttr}, d)$
- (5)  $\text{SelectAttr} = \text{SelectAttr} \cup \{a\}$ ,  $\text{unSelectAttr} = \text{unSelectAttr} - \{a\}$
- (6) If  $\text{EN}(d | \text{SelectAttr}) \neq \text{EN}(d | C)$ , go to Step 2, otherwise  $R = \text{SelectAttr}$
- (7) End

ALGORITHM 2: Depth-first.

$$\begin{aligned} \text{EN}(d | \{X\}) &= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + 3 * \frac{3}{6} \right. \\ &\quad \left. * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] = 2.3774. \end{aligned} \quad (31)$$

Since  $\text{EN}(d | \{X\})$  has the minimum conditional entropy value, we choose the attribute  $X$  as one of selected attributes. Therefore, we can get

$$\begin{aligned} \text{SelectAttr} &= \text{SelectAttr} \cup \{X\} = \{X\}, \\ \text{unSelectAttr} &= \text{unSelectAttr} - \{X\} = \{P, M, S\}. \end{aligned} \quad (32)$$

Since  $\text{EN}(d | \{X\}) \neq \text{EN}(d | C)$ , we still need to add more attributes to  $\text{SelectAttr}$ . On the basis of the rule of heuristic search, we only need to test and compare the conditional entropy values of the subsets in which the condition attribute  $X$  is included:

$$\begin{aligned} \text{EN}(d | \{P, X\}) &= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) \right. \\ &\quad \left. + \frac{3}{6} * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] \end{aligned}$$

$$\begin{aligned} &+ \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \\ &+ \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \Big] = 1.4591, \end{aligned}$$

$$\begin{aligned} \text{EN}(d | \{M, X\}) &= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + 3 * \frac{3}{6} \right. \\ &\quad \left. * \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right] = 2.3774, \end{aligned}$$

$$\begin{aligned} \text{EN}(d | \{S, X\}) &= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + \frac{1}{6} * (1 \log_2 1) \right. \\ &\quad \left. + 2 * \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \right] = 0.6667. \end{aligned} \quad (33)$$

As previously mentioned, since  $\text{EN}(d | \{S, X\})$  is the minimum, we choose the attribute  $S$  as another selected attribute. Thus we have

$$\begin{aligned} \text{SelectAttr} &= \text{SelectAttr} \cup \{S\} = \{S, X\}, \\ \text{unSelectAttr} &= \text{unSelectAttr} - \{S\} = \{P, M\}. \end{aligned} \quad (34)$$

Input: An incomplete decision system  $(U, C \cup \{d\})$  and MaxTries.  
Output: A selected attribute subset  $R$ .

- (1) Initialize SelectAttr =  $C$
- (2) For  $j = 0$  to MaxTries
- (3) Randomly choose a subset of condition attributes  $S_j$
- (4) If  $EN(d | S_j) = EN(d | C)$  and  $|S_j| < |SelectAttr|$
- (5) SelectAttr =  $S_j$
- (6) End
- (7)  $R = SelectAttr$

ALGORITHM 3: LVF.

Since  $EN(d | \{S, X\}) \neq EN(d | C)$ , it is still necessary to choose more attributes to add to the selected attributes in which attributes  $S, X$  are definitely included:

$$\begin{aligned}
& EN(d | \{P, S, X\}) \\
&= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + \frac{1}{6} * (1 \log_2 1) \right. \\
&\quad \left. + \frac{1}{6} * (1 \log_2 1) + \frac{1}{6} * (1 \log_2 1) \right] = 0,
\end{aligned} \tag{35}$$

$$\begin{aligned}
& EN(d | \{M, S, X\}) \\
&= - \left[ 3 * \frac{3}{6} * \left( \frac{3}{3} \log_2 \frac{3}{3} \right) + \frac{1}{6} * (1 \log_2 1) \right. \\
&\quad \left. + 2 * \frac{2}{6} * \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \right] = 0.6667.
\end{aligned}$$

Obviously, since  $EN(d | \{P, S, X\})$  is the minimum, we choose the attribute  $P$ . Thus, we have

$$\begin{aligned}
& SelectAttr = SelectAttr \cup \{P\} = \{P, S, X\}, \\
& unSelectAttr = unSelectAttr - \{P\} = \{M\}.
\end{aligned} \tag{36}$$

And since  $EN(d | \{P, S, X\}) = EN(d | C) = 0$ , the stop criterion is satisfied and the algorithm terminates. Hence, the final result of attributes selection is  $R = \{P, S, X\}$ . The detailed procedure of the Depth-first algorithm is shown in Figure 2.

**4.3. LVF: Probabilistic Search.** Las Vegas algorithm is new for attribute subset selection and can make probabilistic choices of subsets in search of an optimal set. Las Vegas Filter, which is abbreviated as LVF, is a probabilistic algorithm where probabilities of generating any subset are equal [43]. In this paper, we use the investigated conditional entropy as LVF's evaluation measurement. It generates attribute subsets randomly with equal probability and records the minimal size of attributes subset satisfying the stop criterion of maximum tries times. LVF is fast and efficient in reducing the number of candidate features in the early stages and can produce optimal solutions if the computing resources permit. The LVF approach for attribute selection in incomplete decision system is given, as shown in Algorithm 3.

## 5. Experiments

In this section, the performances of our attribute selection algorithms given in previous section are demonstrated and compared. Several real-life incomplete data sets from UCI Repository of Machine Learning Database at the University of California are used in our experiments. These experiments are performed on a personal computer with Windows 7, Intel (R) Core (TM) i3 CPU 2.13 GHz, and 4 GB RAM. The objective of these experiments is to evaluate the effectiveness and efficiency of the previous algorithms. The summary and statistic of the experimental data sets are shown in Table 3 and Figure 3, respectively.

Since some incomplete data sets contain continuous condition attribute values, we conduct a discretization preprocess to turn these continuous values into discrete ones before carrying out attribute selection. The aim of this step is to compress the data and reduce the time consumption of subsequent attribute selection. The running time of each algorithm is average CPU time, expressed in seconds.

In Breadth-first algorithm, we terminate the program when it runs beyond 5000 seconds. Moreover, we set the parameter MaxTries in LVF algorithm to variant values for different incomplete data sets, according to their sizes. The running time and the size of attribute selection results are highly concerned with the choice of the parameter MaxTries. It means that when MaxTries grows, the running time of LVF approach increases linearly and the size of selected attribute subset decreases. Both of them for each attribute selection approach are shown in Table 4. The running time of each approach is the average CPU time, expressed in seconds. We can easily find that Breadth-first approach takes much more time, even more than 5000 seconds, to obtain an attribute reduction, compared with the other two approaches. Three data sets out of six are too large in scale to calculate in limited time. Furthermore, it is also easy to be observed in Table 4 that Depth-first approach tends to select fewer attributes than LVF approach. In other words, the size of attribute subset selected by Depth-first approach intends to be smaller than that of attribute subset selected by LVF approach. And Depth-first approach consumes less time than the other two approaches in most instances. The time consumptions of Breadth-first approach, Depth-first approach, and LVF approach for attribute selection are  $O(2^{|C|} \cdot |U|^2)$ ,  $O(|C|^4 |U|^2)$ , and  $O(\text{MaxTries} \cdot |C| |U|^2)$ , respectively, where MaxTries,  $|C|$ ,

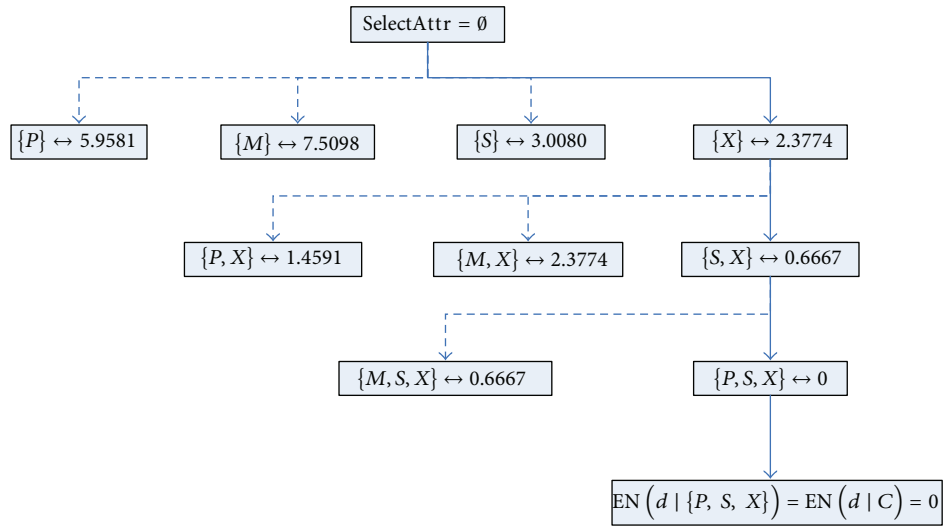


FIGURE 2: The flow chart of the Depth-first algorithm.

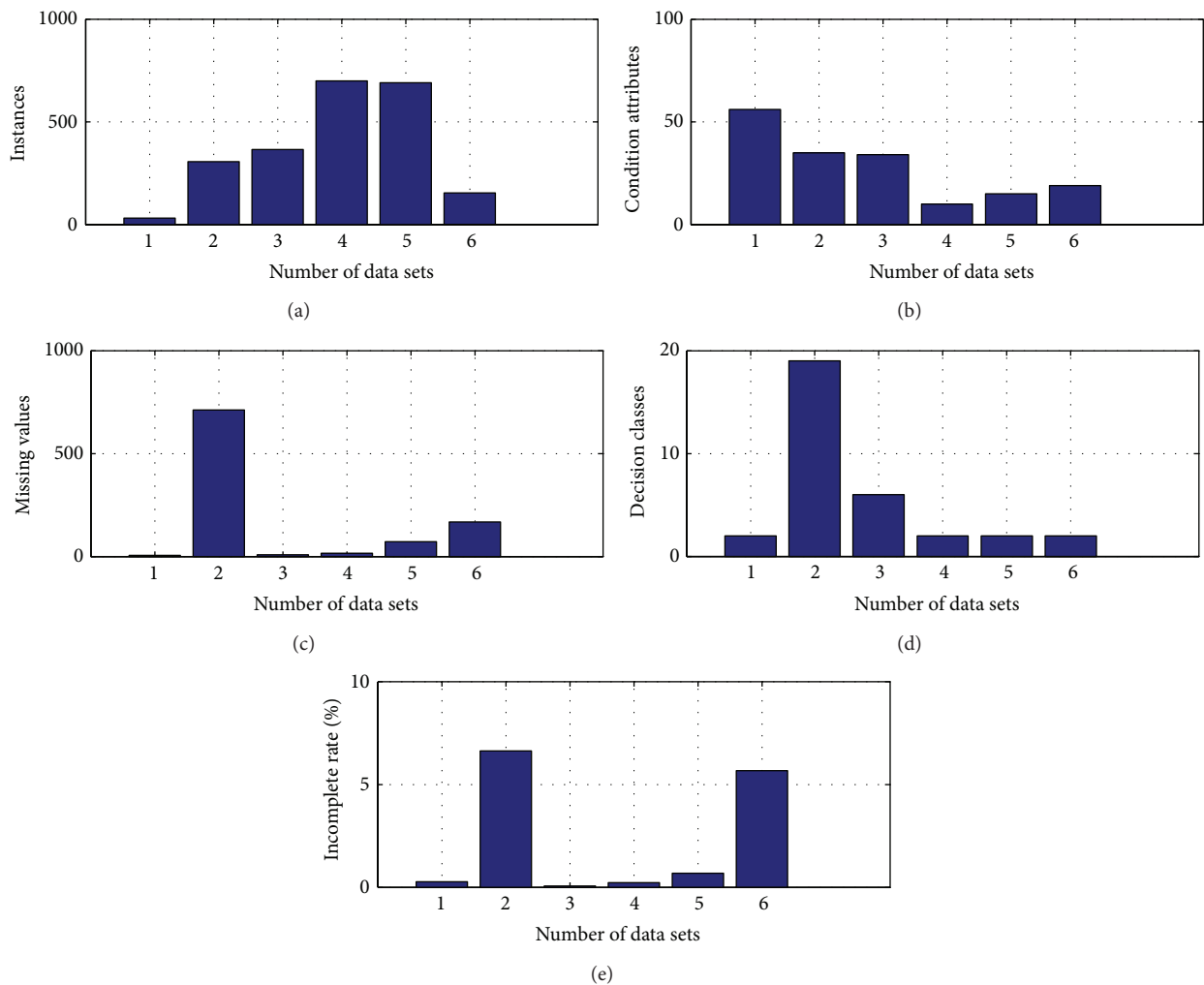


FIGURE 3: The statistical result of experimental data sets.

TABLE 3: The summary of experimental data sets.

Data sets	Instances	Condition attributes	Missing values	Decision classes	Incomplete rate (%)
Lung cancer	32	56	5	3	0.28
Large soybean	307	35	712	19	6.63
Dermatology	366	34	8	6	0.06
Breast cancer	699	10	16	2	0.23
Credit	690	15	71	2	0.69
Hepatitis	155	19	167	2	5.67

TABLE 4: Size of selected attribute subsets and running time.

Data sets	Breadth-first		Depth-first		LVF			Full attr. Size
	Size	Time (s)	Size	Time (s)	Size	Time (s)	MaxTries	
Lung cancer	4	227.59	4	0.68	5	7.34	10000	56
Large soybean	N/A	>5000	14	80.99	23	806.38	5000	35
Dermatology	N/A	>5000	6	58.22	11	44.83	200	34
Breast cancer	5	367.37	5	51.59	5	91.29	100	10
Credit	N/A	>5000	7	89.94	9	90.75	100	15
Hepatitis	5	349.07	6	6.07	6	35.22	500	19

TABLE 5: Performance of attribute selection algorithms with SVM-RBF classifier.

Data sets	Classification accuracy (%)			
	Breadth-first	Depth-first	LVF	Full attr.
Lung cancer	66.92	76.42	69.26	66.67
Large soybean	N/A	95.21	94.74	92.98
Dermatology	N/A	91.22	91.32	83.38
Breast cancer	97.11	97.11	96.18	95.69
Credit	N/A	84.77	81.69	80.88
Hepatitis	83.75	85.00	82.50	78.75

and  $|U|$  denote the MaxTries parameter in LVF approach, the total numbers of the condition attributes, and instances in incomplete data sets, respectively. The relationships between incomplete data sets and the number of selected attributes, the running time of attribute selection, are illustrated in Figures 4 and 5.

The final part of our experiments is to compare and evaluate the efficiency of the proposed algorithms in practical classification tasks. For each of the six data sets in Table 5, we employ the SVM-RBF classifier, which is one of the most frequently used classifiers. We also apply the 10-fold cross-validation method to estimate the classification accuracy with respect to the reducts generated by the proposed algorithms. In each fold, the redundant attributes from the current training set are removed at the beginning, according to the proposed algorithms. Then, the test set is classified by using the rules generated from the training set. The final results of classification accuracies are shown in Table 5. It can be seen that, by using attribute selection algorithms, the classification accuracies for incomplete data sets are all raised in different

degrees, compared with the classification accuracy for the original full attributes. It also can be noticed that the Depth-first algorithm exhibits the highest classification accuracy on each incomplete data set. Therefore, the experimental results demonstrate that the proposed algorithms are effective for attribute selection tasks in application domains.

## 6. Conclusion

In this paper, a rough conditional entropy-based attribute selection approach is proposed to evaluate the significance of condition attributes and find the minimal reduct in incomplete decision systems. By this measure, three types of attribute selection approaches, including the exhaustive search strategy approach Breadth-first, the heuristic search strategy approach Depth-first, and the probabilistic search approach LVF, are constructed. To evaluate the effectiveness of the introduced approaches, experiments on several real-life incomplete data sets are conducted. The experimental results suggest that Depth-first and LVF approaches are practical for attribute selection for classification of high-dimensional data with thousands of condition attributes, and they can efficiently enhance classification accuracy with predominant attributes. However, the process of examining exhaustively all combinations of condition attributes for finding the optimal one is an NP-hard problem. So far, it still cannot be easily calculated by our approaches if there are hundreds of thousands of condition attributes in a complex incomplete decision system. Therefore, for large data sets, to reduce the time consumption of the process of attribute selection, more applicable approaches such as parallel heuristic algorithms are desirable for incomplete decision systems with large scale. This issue needs to be investigated in the future.

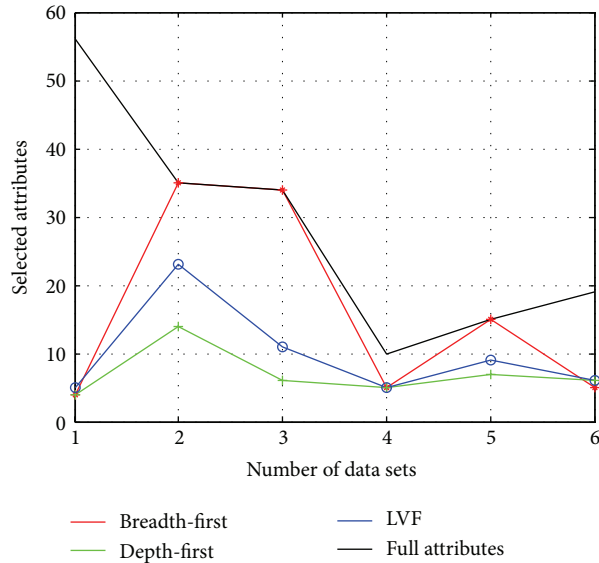


FIGURE 4: Number of selected attributes versus data set.

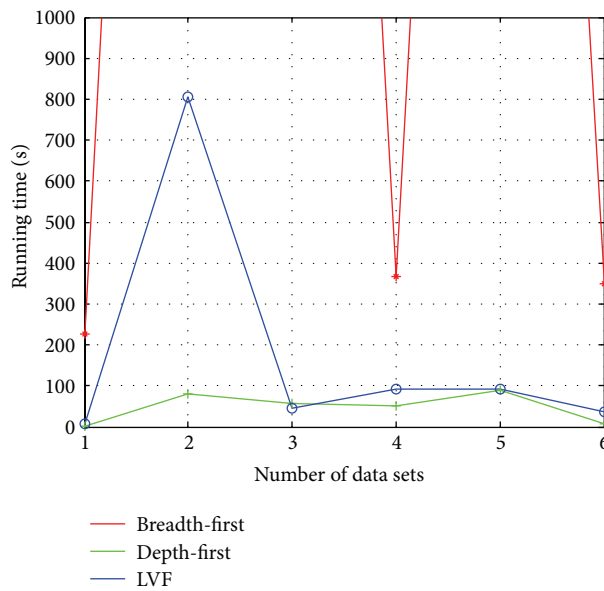


FIGURE 5: Running time of attribute selection versus data set.

**Conflict of Interests**

The authors would like to declare that there is no conflict of interests regarding the publication of the paper.

**Acknowledgments**

The authors would like to thank the anonymous reviewers sincerely for their insightful comments and constructive effort on this investigation. Moreover, the authors thank the UCI Repository of Machine Learning Database at the University of California for providing the experiment data sets. This work is supported by the National Natural Science Foundation of China (no. 61074176).

**References**

- [1] Z. Pawlak, “Rough classification,” *International Journal of Human Computer Studies*, vol. 51, no. 2, pp. 369–383, 1999.
- [2] Z. Pawlak, “Rough classification,” *International Journal of Man-Machine Studies*, vol. 20, no. 5, pp. 469–483, 1984.
- [3] Z. Pawlak, “Rough sets,” *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [4] X. Yi, L. L. Shu, and L. X. Jun, “Generalized rough set model based on set pair connection degree,” in *Proceedings of the 26th Chinese Control Conference (CCC '07)*, pp. 372–375, July 2007.
- [5] X. Zhang and G. Yao, “Generalized rough set model on De Morgan algebras,” in *Proceedings of the IEEE International*

- Conference on Granular Computing (GrC '07)*, pp. 205–208, usa, November 2007.
- [6] S. Trabelsi, Z. Elouedi, and P. Lingras, "Classification systems based on rough sets under the belief function framework," *International Journal of Approximate Reasoning*, vol. 52, no. 9, pp. 1409–1432, 2011.
  - [7] K. Kaneiwa, "A rough set approach to multiple dataset analysis," *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 2538–2547, 2011.
  - [8] L. Sun, C.-Z. Han, N. Dai, and J.-J. Shen, "Feature selection based on Bhattacharyya distance: a generalized rough set method," in *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA '06)*, pp. 10101–10105, June 2006.
  - [9] L. Sun, C. Han, and M. Lei, "Knowledge discovery-based multiple classifier fusion: a generalized rough set method," in *Proceedings of the 9th International Conference on Information Fusion (FUSION '06)*, pp. 1–8, July 2006.
  - [10] Z. Lu and Z. Qin, "Rule extraction from incomplete decision system based on novel dominance relation," in *Proceedings of the 4th International Conference on Intelligent Networks and Intelligent Systems (ICINIS '11)*, pp. 149–152, November 2011.
  - [11] S. Parsons, M. Kubat, and M. Dohnal, "A rough set approach to reasoning under uncertainty," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 7, no. 2, pp. 175–193, 1995.
  - [12] S. Parsons and M. Kubat, "A first-order logic for reasoning under uncertainty using rough sets," *Journal of Intelligent Manufacturing*, vol. 5, no. 4, pp. 211–223, 1994.
  - [13] J. Dai, W. Wang, Q. Xu, and H. Tian, "Uncertainty measurement for interval-valued decision systems based on extended conditional entropy," *Knowledge-Based Systems*, vol. 27, pp. 443–450, 2012.
  - [14] Y. Liu, Y. Chen, and X. Yuan, "Study of metrics system for information fusion evaluation methodology based on rough set theory in intelligent decision-making," in *Proceedings of the International Conference on Management and Service Science (MASS '10)*, pp. 4–10, August 2010.
  - [15] A. Skowron and P. Wasilewski, "Toward interactive Rough-Granular computing," *Control and Cybernetics*, vol. 40, no. 2, pp. 213–235, 2011.
  - [16] A. Skowron, J. Stepaniuk, and R. Swiniarski, "Approximation spaces in rough-granular computing," *Fundamenta Informaticae*, vol. 100, no. 1–4, pp. 141–157, 2010.
  - [17] X. Ye, Y. Tang, Y. Ji, J. Xing, and Y. Bai, "Incomplete information systems based on the set values of attributes," in *Proceedings of the 8th International Conference on Computer Supported Cooperative Work in Design (CSCWD '04)*, pp. 733–738, May 2004.
  - [18] W.-Z. Wu and Y.-H. Xu, "On two types of generalized rough set approximations in incomplete information systems," in *Proceedings of the IEEE International Conference on Granular Computing (GRC '05)*, pp. 303–306, July 2005.
  - [19] G. Liu, "A comparison of two types of generalized rough sets," in *Proceedings of the IEEE International Conference on Granular Computing (GrC '11)*, pp. 423–426, November 2011.
  - [20] I. T. R. Yanto, P. Vitasari, T. Herawan, and M. M. Deris, "Applying variable precision rough set model for clustering student suffering study's anxiety," *Expert Systems with Applications*, vol. 39, no. 1, pp. 452–459, 2012.
  - [21] W.-Z. Wu, "Knowledge acquisition in incomplete information systems based on variable precision rough set model," in *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC '05)*, pp. 2245–2250, August 2005.
  - [22] Z. Min, J.-X. Cheng, and H.-J. Wang, "The research on the classification of the incomplete information system," in *Proceedings of International Conference on Machine Learning and Cybernetics*, pp. 3781–3786, August 2004.
  - [23] M. Kryszkiewicz, "Rules in incomplete information systems," *Information Sciences*, vol. 113, no. 3–4, pp. 271–292, 1999.
  - [24] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4600–4607, 2011.
  - [25] H. Tian and H. Rybinski, "A new approach to computing weighted attributes values in incomplete information systems," in *Proceedings of the International Conference on Hybrid Information Technology (ICHIT '06)*, pp. 268–273, Jeju Island, Korea, November 2006.
  - [26] B. Huang, L. Guo, and X.-Z. Zhou, "Approximation reduction based on similarity relation," in *Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '07)*, pp. 124–128, August 2007.
  - [27] Z. Pawlak and A. Skowron, "Rough sets: some extensions," *Information Sciences*, vol. 177, no. 1, pp. 28–40, 2007.
  - [28] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "An efficient accelerator for attribute reduction from incomplete data in rough set framework," *Pattern Recognition*, vol. 44, no. 8, pp. 1658–1670, 2011.
  - [29] D. Q. Miao, Y. Zhao, Y. Y. Yao, H. X. Li, and F. F. Xu, "Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model," *Information Sciences*, vol. 179, no. 24, pp. 4140–4150, 2009.
  - [30] E. Xu, S.-C. Tong, Y. Wang, S. Xu, and P. Li, "Approach to missing data recovery," in *Proceedings of the International Symposium on Electronic Commerce and Security (ISECS '08)*, pp. 189–192, August 2008.
  - [31] M. Kryszkiewicz, "Rough set approach to incomplete information systems," *Information Sciences*, vol. 112, no. 1–4, pp. 39–49, 1998.
  - [32] B. K. Patra and S. Nandi, "Fast single-link clustering method based on tolerance rough set model," in *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, vol. 5908 of *Lecture Notes in Computer Science*, pp. 414–422, Springer, Berlin, 2009.
  - [33] J. G. Tang, K. She, and F. Zhu, "A new type of covering-based rough fuzzy set model," *Control and Decision*, vol. 27, no. 11, pp. 1653–1662, 2012.
  - [34] J. Tang, K. She, and W. Zhu, "Covering-based rough sets based on the refinement of covering-element," *International Journal of Computational and Mathematical Sciences*, vol. 5, no. 4, pp. 198–208, 2011.
  - [35] B. Huang, H.-X. Li, and D.-K. Wei, "Dominance-based rough set model in intuitionistic fuzzy information systems," *Knowledge-Based Systems*, vol. 28, pp. 115–123, 2012.
  - [36] R. Zhang, B. Liu, and S. Liu, "A Multi-attribute auction model by dominance-based rough sets approach," *Computer Science and Information Systems*, vol. 7, no. 4, pp. 844–858, 2010.
  - [37] M. Kryszkiewicz, "Generalized rules in incomplete information systems," in *Proceedings of 10th International Symposium on Foundations of Intelligent Systems (ISMIS '97)*, pp. 421–430, 1997.



- [38] M. Kryszkiewicz, "Generation of rules from incomplete information systems," *Principles of Data Mining and Knowledge Discovery*, vol. 1263, pp. 156–166, 1997.
- [39] Y. Qian, J. Liang, D. Li, F. Wang, and N. Ma, "Approximation reduction in inconsistent incomplete decision tables," *Knowledge-Based Systems*, vol. 23, no. 5, pp. 427–433, 2010.
- [40] J. Xu and L. Sun, "Knowledge entropy and feature selection in incomplete decision systems," *Applied Mathematics & Information Sciences*, vol. 7, no. 2, pp. 829–837, 2013.
- [41] R. Zhou and E. A. Hansen, "Breadth-first heuristic search," *Artificial Intelligence*, vol. 170, no. 4-5, pp. 385–408, 2006.
- [42] T. A. Akanmu, S. O. Olabiyisi, E. O. Omidiora, C. A. Oyeleye, M. A. Mabayoje, and A. O. Babatunde, "Comparative study of complexities of breadth-first search and depth-first search algorithms using software complexity measures," in *Proceedings of the World Congress on Engineering (WCE '10)*, pp. 203–208, San Francisco, Calif, USA, July 2010.
- [43] G. Nandi, "An enhanced approach to Las Vegas Filter (LVF) feature selection algorithm," in *Proceedings of the 2nd National Conference on Emerging Trends and Applications in Computer Science (NCETACS '11)*, pp. 102–104, March 2011.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

