A Group Incremental Approach to Feature Selection Applying Rough Set Technique

Jiye Liang, Feng Wang, Chuangyin Dang, *Senior Member*, *IEEE*, and Yuhua Qian, *Member*, *IEEE*

Abstract—Many real data increase dynamically in size. This phenomenon occurs in several fields including economics, population studies, and medical research. As an effective and efficient mechanism to deal with such data, incremental technique has been proposed in the literature and attracted much attention, which stimulates the result in this paper. When a group of objects are added to a decision table, we first introduce incremental mechanisms for three representative information entropies and then develop a group incremental rough feature selection algorithm based on information entropy. When multiple objects are added to a decision table, the algorithm aims to find the new feature subset in a much shorter time. Experiments have been carried out on eight UCI data sets and the experimental results show that the algorithm is effective and efficient.

Index Terms—Dynamic data sets, incremental algorithm, feature selection, rough set theory

1 Introduction

It has been observed in many fields that data grow with time in size. This has led to the development of several new analytic techniques. Among these techniques, as an effective and efficient mechanism, incremental approach is often used to discover knowledge from a gradually increasing data set, which can directly carry out the computation using the existing result from the original data set [1], [2], [3], [15], [19], [36], [41]. In recent years, feature selection, as a common technique for data preprocessing in pattern recognition, machine learning, data mining, and so on, has attracted much attention [5], [7], [16], [24]. In this paper, we are concerned with incremental feature selection, which is an extremely important research topic in data mining and knowledge discovery.

On feature selection, a specific theoretical framework is Pawlak's rough set model [13], [31], [45], [53], [54], [55]. Feature selection based on rough set theory is also called attribute reduction [8], [17], [39], [49], [50]. The feature subset obtained by using an attribute reduction algorithm is called a reduct [29], [30]. Attribute reduction is able to select features that preserve the discernibility ability of original ones, but do not attempt to maximize the class separability [14], [18], [26], [40], [47]. In the last two decades, based on rough set theory, many techniques of attribute reduction have been developed [6], [11], [27], [33], [34], [38], [44], [52]. However, most of them can only be applicable to static data tables. When the number of objects increases dynamically in a database, these approaches often need to carry out an attribute reduction algorithm repeatedly and thus consume a huge amount of computational time and memory space. Hence, it is very inefficient to deal with dynamic data tables using these reduction algorithms.

To deal with a dynamically increasing data set, there exists some research on finding reducts in an incremental manner based on rough set theory. Several incremental reduction algorithms have been proposed to deal with dynamic data sets [10], [25], [28], [51]. A common character of these algorithms is that they were only applicable when new data are generated one by one, whereas many real data from applications are generated in groups. When multiple objects are generated at a time in a database, these algorithms may be inefficient since they have to be executed repeatedly to deal with the added group of objects. In other words, when M (e.g., $M = 10{,}000$) objects are generated at a time, one has to execute these algorithms M times. This is obviously very time consuming. If the size of an added object group is very small (e.g., $M = 10$), the existing incremental algorithms may also be effective, of course. However, when massive new objects are generated at a time, this gives rise to much more waste of computational time and space when the existing reduction incremental algorithms are applied. With the development of data processing tools, the speed and volume of data generation increase dramatically. This further appeals for an efficient group incremental attribute reduction algorithm to acquire information timely.

It is well known that the expression of information is usually uncertain and the uncertainties come from disorder, vagueness, approximate expression, and so on. In rough set theory, one of the most common uncertainty measures of data sets is information entropy or its variants. Shannon [37] introduced an entropy to measure the uncertainty of a system, which was called information entropy. Liang et al. [20] introduced a new information entropy called complementary entropy to rough set theory. The complementary

- *J. Liang, F. Wang, and Y. Qian are with Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi Province, China.*
E-mail: ljy@sxu.edu.cn, {sxuwangfeng, jinchengqyh}@126.com.
- *C. Dang is with the Department of System Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. E-mail: mecdang@cityu.edu.hk.*

Manuscript received 3 Oct. 2011; revised 30 Apr. 2012; accepted 8 July 2012; published online 20 July 2012.
Recommended for acceptance by G. Karypis.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-2011-10-0605.
Digital Object Identifier no. 10.1109/TKDE.2012.146.

1041-4347/14/$31.00 © 2014 IEEE Published by the IEEE Computer Society

entropy not only can measure the uncertainty, but also the fuzziness of a rough set. In addition, Qian and Liang [34] proposed another information entropy called combination entropy which can also be used to measure the uncertainty of information systems. As common measures of uncertainty, these three entropies as well as their conditional ones have been widely applied to devise feature selection algorithms [20], [21], [38], [44]. To save the computational time, an accelerator of feature selection was also constructed based on those three entropies in [34]. Although an incremental technique based on the complementary entropy was also reported in [20], it can only be used to update core dynamically.

To fully explore the property of group increments of a data set in feature selection, this paper mainly develops an efficient group incremental reduction algorithm based on the three entropies. In view of that a key step of the development is the computation of entropy, we first introduce in this paper three incremental mechanisms of the three entropies, which determine an entropy by adding objects to a decision table in groups. When a group of objects are added, instead of recomputation on a given data set, the incremental mechanisms derive new entropies by integrating the changes of conditional classes and decision classes into existing entropies. With these mechanisms, a group incremental reduction algorithm is proposed for dynamic decision tables. After a group of objects is added to a decision table, the proposed algorithm generates a reduct for this expanded decision table by fully exploiting the reduct of the original decision table. By doing so, when multiple objects are added to a given decision table, the new reduct can be obtained by the proposed algorithm in a much shorter time. Furthermore, in view of that incremental reduction algorithms based on entropies have not yet been discussed so far, this paper also introduces an incremental reduction algorithm for adding a single object to a decision table. Experiments have been carried out on eight data sets downloaded from UCI. The experimental results show that the proposed algorithm is effective and efficient.

For the convenience of following discussion, here is a description of the main idea in this paper. To select effective features from a dynamically increasing data set, an efficient group incremental feature selection algorithm is proposed in the framework of rough set theory. In the process of selecting useful features, this algorithm employs information entropy to determine feature significance, and significant features are selected as a final feature subset. Experiments show that, compared with both the classical heuristic feature selection algorithms based on information entropy and existing incremental feature selection algorithms, the proposed algorithm can find a feasible feature subset in a much shorter time. The remainder of this paper is organized as follows: Relative works are reviewed in Section 2. Some preliminaries in rough set theory are briefly reviewed in Section 3. Traditional heuristic reduction algorithms based on three representative entropies are introduced in Section 4. Section 5 introduces the incremental feature selection algorithm for adding a single object. And the incremental feature selection algorithm for adding objects in groups is introduced in Section 6. In Section 7, eight UCI data sets are employed to demonstrate the effectiveness and efficiency of the proposed algorithms. Section 8 concludes this paper.

2 Relative Works

In this section, previous research on incremental knowledge updating is reviewed.

Knowledge updating for dynamically increasing data sets has attracted much attention. By integrating the changes of discernibility-matrix, Shan and Ziarko [36] introduced an incremental approach to obtain all maximally generalized rules of a changed decision table. Bang and Zeungnam [2] introduced an incremental learning algorithm to find a minimal set of rules of a decision table. Tong and An [42] constructed the concept of δ-decision matrix, and presented an algorithm for incremental learning of rules. Zheng and Wang [56] developed an effective incremental algorithm which was called RRIA. This algorithm can learn from a domain data set incrementally. Guo et al. [9] proposed an incremental rules extraction algorithm based on the search tree, which is one kind of the first heuristic search algorithms. Furthermore, under variable precision rough-set model (VPRS), Chen et al. [4] introduced a new incremental method for updating approximations of VPRS while objects in the information system dynamically alter.

Feature selection is a common technique for data preprocessing. For incremental feature selection, researchers have also proposed several approaches. Liu [25] proposed an incremental reduction algorithm for the minimal reduct. This algorithm can only be applied to information systems without decision attribute. For decision tables, a reduction algorithm was presented to update reduct in [28], but it was very time consuming. To overcome the deficiencies of these two algorithms, Hu et al. [10] presented an incremental reduction algorithm based on the positive region, and pointed out that this one was more efficient than those two algorithms. Moreover, an incremental reduction algorithm based on the discernibility matrix was proposed by Yang [51].

Rough set theory has been conceived as a powerful soft computing tool to analyze various types of data [29], [30], and is also a specific framework of selecting useful features. Based on rough set theory, to select useful features, a kind of common approaches is using information entropy to measure the feature significance and selecting significant features as a final feature subset [20], [21], [23], [38], [44]. Liang et al. [20], [34] proposed complementary entropy and combination entropy, respectively. These two entropies have been used to determine feature significance in a feature selection algorithm [20], [34]. In [33], information entropy is employed to determine feature significance in an accelerated feature selection algorithm. In [22], Liang et al. proposed an effective feature selection algorithm from a multigranulation view. This algorithm was also designed based on information entropy.

In this paper, to select useful features from a dynamically increasing data set, we focus on incremental feature selection in the framework of rough set theory. In view of that many real data from applications are generated in groups, a group incremental feature selection algorithm is proposed in the framework of rough set theory. And this algorithm employs information entropy to measure the feature significance.

3 Preliminaries on Rough Sets

In this section, several basic concepts in rough set theory are reviewed. In rough set theory, a basic concept is data table, which provides a convenient framework for the representation of records in terms of their attribute values. A data table is a quadruple $S = (U, A, V, f)$, where the universe U is a finite nonempty set of objects (records) and A is a finite nonempty set of attributes (features), $V = \bigcup_{a \in A} V_a$ with V_a being the domain of a, and $f : U \times A \to V$ is an information function with $f(x, a) \in V_a$ for each $a \in A$ and $x \in U$. The table S can often be simplified as $S = (U, A)$.

Each nonempty subset $B \subseteq A$ determines an indiscernibility relation, which is $R_B = \{(x, y) \in U \times U \mid f(x, a) = f(y, a), \forall a \in B\}$. The relation R_B partitions U into some equivalence classes given by $U/R_B = \{[x]_B \mid x \in U\}$, just U/B, where $[x]_B$ denotes the equivalence class determined by x with respect to B, i.e., $[x]_B = \{y \in U \mid (x, y) \in R_B\}$.

Given an equivalence relation R on the universe U and $X \subseteq U$, the lower approximation and upper approximation of X are defined by

$$\underline{R}X = \bigcup \{x \in U \mid [x]_R \subseteq X\},$$

and

$$\overline{R}X = \bigcup \{x \in U \mid [x]_R \cap X \neq \emptyset\},$$

respectively. The order pair $(\underline{R}X, \overline{R}X)$ is called a rough set of X with respect to R. The positive region of X is denoted by $POS_R(X) = \underline{R}X$.

A partial relation $\preceq$ on the family $\{U/B \mid B \subseteq A\}$ is defined as follows: $U/P \preceq U/Q$ (or $U/Q \succeq U/P$) if and only if, for every $P_i \in U/P$, there exists $Q_j \in U/Q$ such that $P_i \subseteq Q_j$, where $U/P = \{P_1, P_2, \ldots, P_m\}$ and $U/Q = \{Q_1, Q_2, \ldots, Q_n\}$ are partitions induced by P, $Q \subseteq A$, respectively. Then, we say that Q is coarser than P, or P is finer than Q. If $U/P \preceq U/Q$ and $U/P \neq U/Q$, we say Q is strictly coarser than P (or P is strictly finer than Q), denoted by $U/P \prec U/Q$ (or $U/Q \succ U/P$). It is clear that $U/P \prec U/Q$ if and only if, for every $X \in U/P$, there exists $Y \in U/Q$ such that $X \subseteq Y$, and there exist $X_0 \in U/P$ and $Y_0 \in U/Q$ such that $X_0 \subset Y_0$.

A decision table is a data table $S = (U, C \cup D)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element of D is called a decision attribute, and D is called a decision attribute set. Given $P \subseteq C$ and $U/D = \{D_1, D_2, \ldots, D_r\}$, the positive region of D with respect to the condition attribute set P is defined by $POS_P(D) = \bigcup_{k=1}^{r} \underline{P}D_k$.

For a decision table S and $P \subseteq C$, $X \in U/P$ is consistent iff all its objects have the same decision value; otherwise, X is inconsistent. A decision table is called a consistent decision table iff all $x \in U$ are consistent; and if $\exists x, y \in U$ are inconsistent, then the table is called an inconsistent decision table. One can extract certain decision rules from a consistent decision table and uncertain decision rules from an inconsistent decision table.

For a decision table S and $P \subseteq C$, when a new object x is added to S, x is indistinguishable on B iff, $\exists y \in U$, $\forall a \in P$, such that $f(x, a) = f(y, a)$; and x is distinguishable on P iff, $\forall y \in U$, $\exists a \in P$ such that $f(x, a) \neq f(y, a)$.

4 Rough Feature Selection Based on Information Entropy

In rough set theory, a given data table usually has multiple reducts, whereas it has been proved that finding its minimal is an NP-hard problem [39]. To overcome this deficiency, researchers have proposed many heuristic reduction algorithms which can generate a single reduct from a given table [11], [12], [20], [21], [33]. Most of these algorithms are of greedy and forward search type. Starting with a nonempty set, these algorithms keep adding one or several attributes of high significance into a pool at each iteration until the dependence no longer increases.

This section reviews the heuristic attribute reduction algorithms based on information entropy for decision tables. The main idea of these algorithms is to keep the conditional entropy of target decision unchanged. This section first reviews three representative entropies, and then introduces the classic attribute reduction algorithm based on information entropy.

In [20], the complementary entropy was introduced to measure uncertainty in rough set theory. Liang and Shi [21] also proposed the conditional complementary entropy to measure uncertainty of a decision table. By preserving the conditional entropy unchanged, the conditional complementary entropy was applied to construct reduction algorithms and reduce the redundant features in a decision table [33]. The conditional complementary entropy used in this algorithm is defined as follows [20], [21], [33].

Definition 1. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Then, one can obtain the partitions $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. Based on these partitions, a conditional entropy of B relative to D is defined as*

$$E(D|B) = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{|X_i \cap Y_j|}{|U|} \frac{|Y_j^c - X_i^c|}{|U|}, \tag{1}$$

where Y_i^c and X_j^c are complement set of Y_i and X_j respectively.

Another information entropy, called combination entropy, was presented in [34] to measure the uncertainty of data tables. The conditional combination entropy was also introduced to construct the heuristic reduction algorithms [34]. This reduction algorithm can find a feature subset that possesses the same number of pairs of indistinguishable elements as that of the original decision table. The definition of the conditional combination entropy is defined as follows [34].

Definition 2. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Then one can obtain the partitions $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. Based on these partitions, a conditional entropy of B relative to D is defined as*

$$CE(D|B) = \sum_{i=1}^{m}\left(\frac{|X_i|}{|U|}\frac{C^2_{|X_i|}}{C^2_{|U|}} - \sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|U|}\frac{C^2_{|X_i\cap Y_j|}}{C^2_{|U|}}\right). \quad (2)$$

where $C^2_{|X_i|}$ denotes the number of pairs of objects which are not distinguishable from each other in the equivalence class X_i.

Based on the classical rough set model, Shannon's information entropy [37] and its conditional entropy were also introduced to find a reduct in a heuristic algorithm [38], [44]. In [44], the reduction algorithm keeps the conditional entropy of target decision unchanged, and the conditional entropy is defined as follows [44].

Definition 3. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Then, one can obtain the partitions $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. Based on these partitions, a conditional entropy of B relative to D is defined as*

$$H(D|B) = -\sum_{i=1}^{m}\frac{|X_i|}{|U|}\sum_{j=1}^{n}\frac{|X_i\cap Y_j|}{|X_i|}\log\left(\frac{|X_i\cap Y_j|}{|X_i|}\right). \quad (3)$$

For convenience, a uniform notation $ME(D|B)$ is introduced to denote the above three entropies. For example, if one adopts Shannon's conditional entropy to define the attribute significance, then $ME(D|B) = H(D|B)$. In [20], [33], and [44], the attribute significance is defined as follows (see Definitions 4 and 5).

Definition 4. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. $\forall a \in B$, the significance measure (inner significance) of a in B is defined as*

$$Sig^{inner}(a, B, D) = ME(D|B - \{a\}) - ME(D|B). \quad (4)$$

Definition 5. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. $\forall a \in C - B$, the significance measure (outer significance) of a in B is defined as*

$$Sig^{outer}(a, B, D) = ME(D|B) - ME(D|B \cup \{a\}). \quad (5)$$

Given a decision table $S = (U, C \cup D)$ and $a \in C$. From the literatures [20], [21], [23], [33], [34], [44], one can get that if $Sig^{inner}(a, C, D) > 0$, then the attribute a is indispensable, i.e., a is a core attribute of S. Based on the core attributes, a heuristic attribute reduction algorithm can find an attribute reduct by gradually adding selected attributes to the core. The definition of reduct based on information entropy is defined as follows [20], [21], [33], [44].

Definition 6. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Then, the attribute set B is a relative reduct if B satisfies:*

1. $ME(D|B) = ME(D|C)$; *and*
2. $\forall a \in B$, $ME(D|B) \neq ME(D|B - \{a\})$.

The first condition guarantees that the reduct has the same distinguish power as the whole attribute set, and the second condition guarantees that there is no redundant attributes in the reduct. Because the heuristic searching strategies in the three algorithms are similar to each other, a common heuristic attribute reduction algorithm based on information entropy for decision tables is introduced as follows [20], [21], [33], [44].

The time complexity of CAR given in [33] is $O(|U||C|^2)$. However, this time complexity does not include the computational time of entropies. Computing entropies is obviously not computationally costless according to the definitions of entropies, and is also a key step in Algorithm 1. To analyze the exact time complexity of Algorithm 1, the time complexity of computing entropies should be given as well.

Algorithm 1. Classic heuristic attribute reduction algorithm based on information entropy for decision tables (CAR)
Input: A decision table $S = (U, C \cup D)$
Output: Reduct red
Step 1: $red \leftarrow \emptyset$;
Step 2: $for(j = 1; j \leq |C|; j++)$
{If $Sig^{inner}(a_j, C, D) > 0$, then $red \leftarrow red \cup \{a_j\}$;}
Step 4: $P \leftarrow red$, $while(ME(D|P) \neq ME(D|C))$ do
{Compute and select sequentially
$Sig^{outer}(a_0, P, D) = max\{Sig^{outer}(a_i, P, D),$
$a_i \in C - P\}$;
$P \leftarrow P \cup \{a_0\}$;}
Step 5: $red \leftarrow P$, return red and end.

According to Definitions 1-3, a decision table first needs to compute its conditional classes and decision classes, and then computes its value of entropy. Xu et al. [48] gave a fast algorithm for partition with time complexity being $O(|U||C|)$. So, the time complexity of computing entropy is $O(|U||C| + |U| + \sum_{i=1}^{m}|X_i| \cdot \sum_{j=1}^{n}|Y_j|) = O(|U|^2)$ (the specific introduction of m, n, X_i and Y_j is shown in Definitions 1-3). Thus, the time complexity of computing core (Steps 1 and 2) is $O(|C||U|^2)$, and the time complexity of computing reduct according to CAR is $O(|C||U|^2 + |C|(|U||C| + |U|^2)) = O(|C|^2|U| + |C||U|^2)$.

5 Incremental Feature Selection Algorithm for Adding a Single Object

Given a dynamic decision table, based on those three representative entropies, this section introduces an incremental feature selection algorithm for adding a single object. This section is divided into two parts. Section 5.1 introduces the incremental mechanisms for the three entropies. When a new object is added to a given decision table, instead of recomputation on the new decision table, the incremental mechanisms aim to calculate new entropies by integrating the changes of classes into the existing entropies of the original decision table. Section 5.2 introduces the incremental feature selection algorithm based on information entropy for adding a single object. Similarly, this incremental algorithm finds a new feature subset on the available result of feature selection. The incremental mechanisms of entropies are used in the steps of the algorithm where entropies are computed. To make the presentation easier to follow, some illustrative examples are also given in this section.

5.1 Incremental Mechanism to Calculate Entropies After Adding a Single Object

Given a dynamic decision table, with the increase of objects, recomputing entropy is obviously time consuming. To address this issue, this section introduces three incremental mechanisms for computing entropies. When a single object is added to a decision table, Theorems 1-4 introduce the incremental mechanisms for the three entropies respectively.

In [23], when a single object is added to a given decision table, the incremental mechanism of complementary conditional entropy (see Definition 1) has been analyzed, which is shown in Theorem 1.

Theorem 1. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The complementary conditional entropy of D with respect to B is $E_U(D|B)$. Suppose that object x is added to the table S, $x \in X'_p$ and $x \in Y'_q$ ($X'_p \in U \cup \{x\}/B$ and $Y'_q \in U \cup \{x\}/D$). Then, the new complementary conditional entropy becomes*

$$E_{U\cup\{x\}}(D|B) = \frac{1}{(|U|+1)^2}(|U|^2 E_U(D|B) + 2|X'_p - Y'_q|).$$

Proof. The proof can be found in [23]. □

For the convenience of introducing incremental mechanism of combination entropy, here gives a variant of the definition of combination entropy (see Definition 2). According to $C_N^2 = \frac{N(N-1)}{2}$, Definition 7 shows a variant of combination entropy. Based on this variant, the incremental mechanism of combination entropy is introduced in Theorem 2.

Definition 7. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. One can obtain the condition partition $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The conditional entropy of B relative to D is defined as*

$$CE(D|B) = \sum_{i=1}^{m}\left(\frac{|X_i|^2(|X_i|-1)}{|U|^2(|U|-1)} - \sum_{j=1}^{n}\frac{|X_i \cap Y_j|^2(|X_i \cap Y_j|-1)}{|U|^2(|U|-1)}\right). \quad (6)$$

Theorem 2. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$, and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The conditional combination entropy of D with respect to B is $CE_U(D|B)$. Suppose that a new object x is added to the table S, $x \in X'_p$ and $x \in Y'_q$ ($X'_p \in U \cup \{x\}/B$ and $Y'_q \in U \cup \{x\}/D$). Then, the new combination conditional entropy becomes*

$$CE_{U\cup\{x\}}(D|B) = \frac{1}{(|U|+1)^2}(|U|(|U|-1)CE_U(D|B) + |X'_p - Y'_q|(3|X'_p| + 3|X'_p \cap Y'_q| - 5)).$$

The following two theorems are the introduction of incremental mechanism of Shannon's information entropy (see Definition 3).

Theorem 3. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The Shannon's conditional entropy of D with respect to B is $H_U(D|B)$. Suppose that a new object x is added to the table S, $x \in X'_p$ and $x \in Y'_q$ ($X'_p \in U \cup \{x\}/B$ and $Y'_q \in U \cup \{x\}/D$). The new Shannon's conditional entropy becomes*

$$H_{U\cup\{x\}}(D|B) = \frac{1}{(|U|+1)}(|U|H_U(D|B) - \Delta),$$

where

$$\Delta = \sum_{j=1}^{n-1}|(X'_p - \{x\}) \cap Y_j| log\frac{|X'_p|-1}{|X'_p|} + (|X'_p \cap Y'_q| - 1) log\frac{(|X'_p|-1)|X'_p \cap Y'_q|}{|X'_p|(|X'_p \cap Y'_q|-1)} + log\frac{|X'_p \cap Y'_q|}{|X'_p|}.$$

TABLE 1
A Decision Table

	c_1	c_2	c_3	c_4	d
x_1	1	0	1	0	0
x_2	1	0	1	0	1
x_3	1	1	0	0	0
x_4	0	1	0	0	1

Obviously, the Δ in Theorem 3 is relatively complicated, which may give rise to much waste of computational time, especially for large-scale data sets. Thus, Theorem 4 shows an approximate computational formula.

Theorem 4. *Let $S = (U, C \cup D)$ be a large-scale decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$, and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The Shannon's conditional entropy of D with respect to B is $H_U(D|B)$. Suppose that a new object x is added to the table S, $x \in X'_p$ and $x \in Y'_q$ ($X'_p \in U \cup \{x\}/B$ and $Y'_q \in U \cup \{x\}/D$). The new Shannon's conditional entropy becomes*

$$H_{U\cup\{x\}}(D|B) \approx \frac{1}{(|U|+1)}\left(|U|H_U(D|B) - log\frac{|X'_p \cap Y'_q|}{|X'_p|}\right).$$

In the following, we employ an example to illustrate above incremental mechanisms.

Example 1. Let Table 1 be a decision table. In this table, $U = \{x_1, x_2, x_3, x_4\}$ is the universe, $C = \{c_1, c_2, c_3, c_4\}$ is the condition attribute set and $D = \{d\}$ is the decision attribute.

We have that $U/C = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$ and $U/D = \{\{x_1, x_3\}, \{x_2, x_4\}\}$.

According to Definitions 1-3 (or 1, 3, and 7), we have that $E_U(D|C) = \frac{1}{8}$, $CE_U(D|C) = \frac{1}{12}$, and $H_U(D|C) \approx 0.15$.

Suppose that new object $x_5 = \{1, 0, 1, 1, 1\}$ is added to Table 1. We have $X'_p = \{x_5\}$ and $Y'_q = \{x_2, x_4, x_5\}$.

Then, according to Theorem 1, we have $|X'_p - Y'_q| = |\{x_5\} - \{x_2, x_4, x_5\}| = 0$ and $E_{U\cup\{x\}}(D|B) = \frac{1}{(4+1)^2}(4^2 \times \frac{1}{8} + 2 \times 0) = 0.08$.

According to Theorem 2, we have $|X'_p| = 1$, $|X'_p \cap Y'_q| = 1$, and $|X'_p - Y'_q| = 0$. Thus, $CE_{U\cup\{x\}}(D|B) = \frac{1}{(4+1)^2}(4 \times (4-1) \times \frac{1}{12} + 0 \times (3 \times 1 + 3 \times 1 - 5)) = 0.04$.

According to Theorem 3, we have $|X'_p| = 1$ and $|X'_p \cap Y'_q| = 1$. Thus, $H_{U\cup\{x\}}(D|B) = \frac{1}{(4+1)}(4 \times 0.15 - 0) = 0.12$.

Because the size of Table 1 employed in this example is very small, we used Theorem 3 to compute Shannon's entropy. For the larger data sets employed in the section of experiments, Theorem 4 is used to compute entropy.

5.2 Incremental Algorithm for Adding a Single Object

Based on incremental mechanisms of the three entropies, this section introduces an incremental feature selection algorithm based on information entropy in the framework of rough set theory.

Given a decision table $S = (U, C \cup D)$. Suppose that $B \subseteq C$ is a reduct of S and x is the new incremental object. There are three distinguishing situations about x based on the reduct B:

1. x is distinguishable on B, and x is also distinguishable on C;
2. x is indistinguishable on B, and x is distinguishable on C; and
3. x is indistinguishable on B, and x is also indistinguishable on C.

For above three distinguishing situations, following three theorems introduce the changes of the three entropies.

Theorem 5. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Supposed that B is a reduct of S and x is a new incremental object. Then, if x is distinguishable on both B and C, then $ME_{U\cup\{x\}}(D|B) = ME_{U\cup\{x\}}(D|C)$.*

Theorem 6. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Supposed that B is a reduct of S and x is a new incremental object. Then, if x is indistinguishable on B and is distinguishable on C, then $ME_{U\cup\{x\}}(D|B) \neq ME_{U\cup\{x\}}(D|C)$.*

Theorem 7. *Let $S = (U, C \cup D)$ be a decision table and $B \subseteq C$. Supposed that B is a reduct of S and x is a new incremental object. Then, if x is indistinguishable both on B and C, then $ME_{U\cup\{x\}}(D|B) = ME_{U\cup\{x\}}(D|C)$.*

According to Theorems 5 and 7, if the added object is distinguishable or indistinguishable on both B and C, then new entropies of D with respect to B and C are identical. Hence, according to the definition of reduct (see Definition 6), it only need to delete the reductant attributes from B for these two situations. If the added object is indistinguishable on the previous reduct B and is distinguishable on conditional attribute set C, finding new reduct needs to add new attributes. On this basis, Algorithm 3 introduces an incremental algorithm for reduct computation.

An example is employed to illustrate Algorithm 2. For convenience, Example 2 shows the process of computing reduct based on complementary entropy. In the same way, one can compute reduct based on the other two entropies by using Algorithm 2.

Algorithm 2. An incremental algorithm for reduct computation ($IARC$)

Input: A decision table $S = (U, C \cup D)$, reduct RED_U on U, and the new incremental object x

Output: Attribute reduct $RED_{U\cup\{x\}}$ on $U \cup \{x\}$

Step 1 : $B \leftarrow RED_U$. Find M'_t: in $U/B = \{M_1, M_2, \ldots, M_l\}$, if all of the attribute values of x is identical to that of M_t on B, then $M'_t = M_t \cup \{x\}$; else $M'_t = \{x\}$.

Step 2 : If $M'_t = \{x\}$, then turn to *Step 5*; if $M'_t = M_t \cup \{x\}$, then turn to *Step 3*.

Step 3 : Find X'_p: similarly, in $U/C = \{X_1, X_2, \ldots, X_m\}$, if $X'_p = X_p \cup \{x\}$, then turn to *Step 5*; if $X'_p = \{x\}$, then turn to *Step 4*.

Step 4 : While $ME_{U\cup\{x\}}(D|B) \neq ME_{U\cup\{x\}}(D|C)$ *do*
{For each $a \in C - B$, compute $Sig^{outer}_{U\cup\{x\}}(a, B, D)$ (according to Theorems 1, 2, or 4 and Definition 5);
Select $a_0 = max\{Sig^{outer}_{U\cup\{x\}}(a, B, D)\}$, $a \in C - B$;
$B \leftarrow B \cup \{a_0\}$.}

Step 5 : For each $a \in B$ *do*
{Compute $Sig^{inner}_{U\cup\{x\}}(a, B, D)$;
If $Sig^{inner}_{U\cup\{x\}}(a, B, D) = 0$, then $B \leftarrow B - \{a\}$.}

Step 6 : $RED_{U\cup\{x\}} \leftarrow B$, return $RED_{U\cup\{x\}}$ and end.

Algorithm 3. A group incremental algorithm for reduct computation ($GIARC$)

Input: A decision table $S = (U, C \cup D)$, reduct RED_U on U, and the new object set U_X

Output: Reduct $RED_{U\cup U_X}$ on $U \cup U_X$

Step 1 : $B \leftarrow RED_U$. Compute $U/B = \{X^B_1, X^B_2, \ldots, X^B_m\}$, $U/C = \{X^C_1, X^C_2, \ldots, X^C_s\}$, $U_X/B = \{M^B_1, M^B_2, \ldots, M^B_{m'}\}$, and $U_X/C = \{M^C_1, M^C_2, \ldots, M^C_{s'}\}$.

Step 2 : Compute $(U \cup U_X)/B = \{X'^B_1, X'^B_2, \ldots, X'^B_k, X^B_{k+1}, X^B_{k+2}, \ldots, X^B_m, M^B_{k+1}, M^B_{k+2}, \ldots, M^B_{m'}\}$ and $(U \cup U_X)/C = \{X'^C_1, X'^C_2, \ldots, X'^C_{k'}, X^C_{k'+1}, X^C_{k'+2}, \ldots, X^C_s, M^C_{k'+1}, M^C_{k'+2}, \ldots, M^C_{s'}\}$.

Step 3 : If $k = 0$ and $k' = 0$, turn to *Step 4*; else turn *Step 5*.

Step 4 : Compute $ME_{U_X}(D|B)$ and $ME_{U_X}(D|C)$.
If $ME_{U_X}(D|B) = ME_{U_X}(D|C)$, turn to *Step 7*;
else turn to *Step 5*.

Step 5 : *while* $ME_{U\cup U_X}(D|B) \neq ME_{U\cup U_X}(D|C)$ *do*
{For each $a \in C - B$, compute $Sig^{outer}_{U\cup U_X}(a, B, D)$;
Select $a_0 = max\{Sig^{outer}_{U\cup U_X}(a, B, D)$, $a \in C - B\}$;
$B \leftarrow B \cup \{a_0\}$.
}

Step 6 : For each $a \in B$ *do*
{Compute $Sig^{inner}_{U\cup U_X}(a, B, D)$;
If $Sig^{inner}_{U\cup U_X}(a, B, D) = 0$, then $B \leftarrow B - \{a\}$.
}

Step 7 : $RED_{U\cup U_X} \leftarrow B$, return $RED_{U\cup U_X}$ and end.

Example 2 (Continued from Example 1). Computing new reduct based on complementary entropy by using Algorithm 2.

For Table 1, its previous reduct found by using Algorithm 1 based on complementary entropy is $\{c_1, c_2\}$. Suppose that new object $x_5 = \{1, 0, 1, 1, 1\}$ is added to Table 1.

According to Step 1, we have $M'_t = \{x_1, x_2, x_5\}$. Obviously, $M'_t \neq \{x_5\}$, then algorithm turns to Step 3 according to Step 2.

According to Step 3, we have $X'_p = \{x_5\}$. Hence, algorithm turns to Step 4 according to Step 3.

From Theorem 1, one can get $E_{U\cup\{x\}}(D|B) = 0.16$, $E_{U\cup\{x\}}(D|C) = 0.08$, and $E_{U\cup\{x\}}(D|B) \neq E_{U\cup\{x\}}(D|C)$. Thus, algorithm needs to add attributes from $C - B$ according to Step 4.

In the first circulation, $Sig^{outer}_{U\cup\{x\}}(c_3, B, D) = 0$ and $Sig^{outer}_{U\cup\{x\}}(c_4, B, D) = 0.08$. Then, we have $B = \{c_1, c_2\} \cup \{c_4\} = \{c_1, c_2, c_4\}$. Now, we have $E_{U\cup\{x\}}(D|B) = 0.16$ and $E_{U\cup\{x\}}(D|B) = E_{U\cup\{x\}}(D|C)$. Algorithm here stop the circulation in Step 4.

TABLE 2
Comparison of Time Complexity

Entropy	Classic	Incremental
	$O(\|U\|^2)$	$O(\|U\|\|C\| + \|X_p'\|\|Y_q'\|)$
Reduct	CAR	$IARC$
	$O(\|C\|^2\|U\| + \|C\|\|U\|^2)$	$O(\|C\|^2\|U\| + \|C\|\|X_p'\|\|Y_q'\|)$

According to Step 5, there is no attribute in B need to be deleted. Thus, $RED_{U\cup\{x\}} \leftarrow B$ and $RED_{U\cup\{x\}} = \{c_1, c_2, c_4\}$.

The following is the time complexities of Algorithm 2. Here are some explanations first. Based on the analysis in Section 5.1, when x is added to the table, one can also get the new value of entropy by using the incremental formulas. And the time complexity of computing entropy is $O(|U||C| + |U| + m|C| + n + |X_p'||Y_q'|) = O(|U||C| + |X_p'||Y_q'|)$ (the explanations of m, n, X_p' and Y_q' are shown in Theorems 1, 2, and 4). For convenience, we make Θ' to denote the above time complexity, i.e., $\Theta' = O(|U||C| + |X_p'||Y_q'|)$.

In the algorithm $IARC$, the time complexity of Steps 1 and 3 is $O(|U||C|)$. In Step 4, the time complexity of adding attributes is $O(|C|\Theta')$. In Step 5, the time complexity of deleting redundant attributes is $O(|C|\Theta')$. Hence, the total time complexity of algorithm $IARC$ is $O(|U||C| + |C|(|U||C| + |X_p'||Y_q'|)) = O(|U||C|^2 + |C||X_p'||Y_q'|)$. To stress the above findings, Table 2 shows the time complexities of computing reduct.

From Table 2, because of that $|X_p'||Y_q'|$ is usually much smaller than $|U|^2$, we can conclude that the computational time of new incremental algorithms are usually much smaller than that of the classic algorithms. Note that, sometimes, $|X_p'||Y_q'|$ may be identical to $|U|^2$, i.e., $|X_p'| = |U|$ and $|Y_q'| = |U|$. In this situation, the discernibility ability of the attributes induced $|X_p'|$ (or $|Y_q'|$) is very weaker, and thus these attributes will have few contributions to select effective feature subset. In other words, it is impossible that these attributes can be selected as useful features. Hence, $|X_p'||Y_q'|$ is more commonly much smaller than $|U|^2$ in the process of selecting effective features, and the new incremental algorithms can save more computation than CAR.

6 Incremental Feature Selection Algorithm for Adding Multiple Objects

In practice, the rapid development of data processing tools has led to the high speed of dynamic data updating. Thus, many real data in applications may be generated in groups instead of one by one. If multiple objects are added to databases, the feature selection algorithm proposed in the previous section may be less efficient. In other words, the incremental algorithm for single object needs to be reperformed repeatedly to deal with multiple objects. This obviously gives rise to much waste of computational time. To overcome this deficiency, this section introduces a group incremental feature selection algorithm, which aims to deal with multiple objects at a time instead of repeatedly.

This section is divided into two parts to introduce the group incremental algorithm. We assume in this paper that the size of an added object set is smaller than that of the original table. Section 6.1 introduces the incremental mechanisms of three entropies for adding multiple objects. When multiple objects are added to a given decision table, the incremental mechanisms aim to compute new entropy by using the previous entropy instead of recomputation on the decision table. Section 6.2 introduces the group incremental feature selection algorithm based on information entropy. The incremental mechanisms of entropies are used in the steps of the algorithm which need to compute entropy. To make the presentation easier to follow, some examples are also given in this section.

6.1 Incremental Mechanism to Calculate Entropies After Adding Multiple Objects

Given a decision table, when multiple objects are added, the incremental mechanisms introduced in Section 5.1 for computing entropy obviously need to repeat the operation many times. Hence, this section introduces the group incremental mechanisms of entropies. Theorems 8-10 introduce the group incremental mechanisms of three entropies, respectively.

For convenience, here are some explanations which will be used in the following theorems. Given a decision table $S = (U, C \cup D), B \subseteq C, U/B = \{X_1, X_2, \ldots, X_m\}$, and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. Suppose that U_X is the incremental object set, $U_X/B = \{M_1, M_2, \ldots, M_{m'}\}$ and $U_X/D = \{Z_1, Z_2, \ldots, Z_{n'}\}$. In the view of that, between U/B and U_X/B, there may be some conditional classes with the identical attribute values on B, we might as well assume that $(U \cup U_X)/B = \{X_1', X_2', \ldots, X_k', X_{k+1}, X_{k+2}, \ldots, X_m, M_{k+1}, M_{k+2}, \ldots, M_{m'}\}$ and $(U \cup U_X)/D = \{Y_1', Y_2', \ldots, Y_l', Y_{l+1}, Y_{l+2}, \ldots, Y_n, Z_{l+1}, Z_{l+2}, \ldots, Z_{n'}\}$. In $(U \cup U_X)/B$, $X_i' = X_i \cup M_i (i = 1, 2, \ldots, k)$ denote the combinative conditional classes, that is, the attribute values of $X_i \in U/B$ and $M_i \in U_X/B$ are identical. And $X_i \in U/B (i = k+1, 2, \ldots, m)$ and $M_j \in U_X/B (j = k+1, k+2, \ldots, m')$ denote the conditional classes which cannot be combined. Similarly, in $(U \cup U_X)/D$, $Y_i' = Y_i \cup Z_i (i = 1, 2, \ldots, l)$ denote the combinative of decision classes with the identical attribute values on D. And $Y_i \in U/D (i = l+1, l+2, \ldots, n)$ and $Z_j \in U_X/D (j = l+1, l+2, \ldots, n')$ denote the decision classes which can not be combined.

Example 3. Let $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, $U/B = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5\}, \{x_6, x_7\}\}$ and $U/D = \{\{x_1, x_2, x_3\}, \{x_4\}, \{x_5\}, \{x_6, x_7\}\}$. The incremental data set $U_X = \{y_1, y_2, y_3, y_4\}$, $U_X/B = \{\{y_1, y_2\}, \{y_3\}\}$, and $U_X/D = \{\{y_1\}, \{y_2, y_3\}\}$.

It is assumed that the attribute values of $\{y_3\}$ is identical to that of $\{x_5\}$ with respect to B, and the decision attribute value of $\{y_2, y_3\}$ is identical to that of $\{x_6, x_7\}$. Then, one have

$$(U \cup U_X)/B = \{\{x_5, y_3\}, \{x_1, x_2\}, \{x_3, x_4\}, \{x_6, x_7\}, \{y_1, y_2\}\},$$

where, $X_1' = \{x_5, y_3\}$, $X_2 = \{x_1, x_2\}$, $X_3 = \{x_3, x_4\}$, $X_4 = \{x_6, x_7\}$, and $M_2 = \{y_1, y_2\}$.

$$(U \cup U_X)/D = \{\{x_6, x_7, y_2, y_3\}, \{x_1, x_2, x_3\}, \{x_4\}, \{x_5\}, \{y_1\}\},$$

where, $Y_1' = \{x_6, x_7, y_2, y_3\}$, $Y_2 = \{x_1, x_2, x_3\}$, $Y_3 = \{x_4\}$, $Y_4 = \{x_5\}$, and $Z_2 = \{y_1\}$. Obviously, $m = 4, n = 4$, $m' = 2, n' = 2, k = 1$, and $l = 1$.

Given a decision table, Theorem 8 introduces the incremental mechanism based on complementary entropy.

Theorem 8. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$, and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The complementary conditional entropy of D with respect to B is $E_U(D|B)$. Suppose that U_X is an incremental object set, $U_X/B = \{M_1, M_2, \ldots, M_{m'}\}$ and $U_X/D = \{Z_1, Z_2, \ldots, Z_{n'}\}$. We assume that $(U \cup U_X)/B = \{X_1', X_2', \ldots, X_k', X_{k+1}, X_{k+2}, \ldots, X_m, M_{k+1}, M_{k+2}, \ldots, M_{m'}\}$ and $(U \cup U_X)/D = \{Y_1', Y_2', \ldots, Y_l', Y_{l+1}, Y_{l+2}, \ldots, Y_n, Z_{l+1}, Z_{l+2}, \ldots, Z_{n'}\}$. Then, the new complementary conditional entropy becomes*

$$E_{U \cup U_X}(D|B) = \frac{1}{(|U \cup U_X|)^2}(|U|^2 E_U(D|B) + |U_X|^2 E_{U_X}(D|B)) + \Delta,$$

where

$$\Delta = \sum_{i=1}^{k}\left(\sum_{j=1}^{l}\frac{|X_i \cap Y_j||M_i - Z_j| + |M_i \cap Z_j||X_i - Y_j|}{(|U \cup U_X|)^2} + \sum_{j=l+1}^{n}\frac{|X_i \cap Y_j||M_i|}{(|U \cup U_X|)^2} + \sum_{j=l+1}^{n'}\frac{|M_i \cap Z_j||X_i|}{(|U \cup U_X|)^2}\right).$$

In what following, the group incremental mechanism based on combination entropy is introduced in Theorem 9.

Theorem 9. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The combination conditional entropy of D with respect to B is $CE_U(D|B)$. Suppose that U_X is an incremental object set, $U_X/B = \{M_1, M_2, \ldots, M_{m'}\}$ and $U_X/D = \{Z_1, Z_2, \ldots, Z_{n'}\}$. We assume that $(U \cup U_X)/B = \{X_1', X_2', \ldots, X_k', X_{k+1}, X_{k+2}, \ldots, X_m, M_{k+1}, M_{k+2}, \ldots, M_{m'}\}$ and $(U \cup U_X)/D = \{Y_1', Y_2', \ldots, Y_l', Y_{l+1}, Y_{l+2}, \ldots, Y_n, Z_{l+1}, Z_{l+2}, \ldots, Z_{n'}\}$. Then, the new combination conditional entropy becomes*

$$CE_{U \cup U_X}(D|B) = \frac{1}{(|U| + |U_X|)^2(|U| + |U_X| - 1)}(|U|^2(|U| - 1)CE_U(D|B) + |U_X|^2(|U_X| - 1)CE_{U_X}(D|B)) + \Delta,$$

where

$$\Delta = \sum_{i=1}^{k}\left(\frac{|X_i||M_i|(3|X_i| + 3|M_i| - 2)}{(|U| + |U_X|)^2(|U| + |U_X| - 1)} - \sum_{j=1}^{l}\frac{|X_i \cap Y_j||M_i \cap Z_j|(3|X_i \cap Y_j| + 3|M_i \cap Z_j| - 2)}{(|U| + |U_X|)^2(|U| + |U_X| - 1)}\right).$$

Based on Shannon's entropy, the group incremental mechanism for adding multiple objects is introduced in Theorem 10.

Theorem 10. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$, $U/B = \{X_1, X_2, \ldots, X_m\}$, and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. The Shannon's conditional entropy of D with respect to B is $H_U(D|B)$. Suppose that U_X is an incremental object set, $U_X/B = \{M_1, M_2, \ldots, M_{m'}\}$ and $U_X/D = \{Z_1, Z_2, \ldots, Z_{n'}\}$. We assume that $(U \cup U_X)/B = \{X_1', X_2', \ldots, X_k', X_{k+1}, X_{k+2}, \ldots, X_m, M_{k+1}, M_{k+2}, \ldots, M_{m'}\}$ and $(U \cup U_X)/D = \{Y_1', Y_2', \ldots, Y_l', Y_{l+1}, Y_{l+2}, \ldots, Y_n, Z_{l+1}, Z_{l+2}, \ldots, Z_{n'}\}$. Then, the new Shannon's conditional entropy becomes*

$$H_{U \cup U_X}(D|B) = \frac{1}{|U| + |U_X|}(|U|H_U(D|B) + |U_X|H_{U_X}(D|B)) - \Delta,$$

where

$$\Delta = \sum_{i=1}^{k}\left(\sum_{j=1}^{l}\left(\frac{|X_i \cap Y_j|}{|U| + |U_X|}log\frac{|X_i||X_i' \cap Y_j'|}{|X_i'||X_i \cap Y_j|} + \frac{|M_i \cap Z_j|}{|U| + |U_X|}log\frac{|M_i||X_i' \cap Y_j'|}{|X_i'||M_i \cap Z_j|}\right) + \sum_{j=l+1}^{n}\frac{|X_i \cap Y_j|}{|U| + |U_X|}log\frac{|X_i|}{|X_i'|} + \sum_{j=l+1}^{n'}\frac{|M_i \cap Z_j|}{|U| + |U_X|}log\frac{|M_i|}{|X_i'|}\right).$$

To illustrate above study clearly, here employs an example to introduce the process of computing entropies in a group incremental way.

Example 4. For Table 1, suppose that $U_X = \{x_5, x_6, x_7\}$ is the added object set. $x_5 = \{1, 0, 1, 1, 1\}$, $x_6 = \{0, 1, 0, 0, 0\}$, and $x_7 = \{1, 1, 0, 0, 0\}$.

We have that $U/C = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$, $U/D = \{\{x_1, x_3\}, \{x_2, x_4\}\}$, $U_X/C = \{\{x_5\}, \{x_6\}, \{x_7\}\}$, and $U_X/D = \{\{x_5\}, \{x_6, x_7\}\}$.

Then, one can get that $U \cup U_X/C = \{\{x_3, x_7\}, \{x_4, x_6\}, \{x_1, x_2\}, \{x_5\}\}$ and $U \cup U_X/D = \{\{x_1, x_3, x_6, x_7\}, \{x_2, x_4, x_5\}\}$.

According to Definitions 1-3, we have that $E_U(D|C) = \frac{1}{8}$, $CE_U(D|C) = \frac{1}{12}$, $H_U(D|C) \approx 0.15$, and $E_{U \cup U_X}(D|C) = CE_{U \cup U_X}(D|C) = H_{U \cup U_X}(D|C) = 0$.

According to Theorem 8, we have that $k = 2, m = 3$, $m' = 3, l = 2, n = 2$, and $n' = 2$. And $X_1' = \{x_3, x_7\}$, $X_2' = \{x_4, x_6\}$, $X_3 = \{x_1, x_2\}$, and $M_3 = \{x_5\}$. $Y_1' = \{x_1, x_3, x_6, x_7\}$ and $Y_2' = \{x_2, x_4, x_5\}$. Hence, $E_{U \cup U_X}(D|C) = \frac{1}{7^2} \times (4^2 \times \frac{1}{8} + 3^2 \times 0) + \frac{2}{7^2} = \frac{2}{49}$.

TABLE 3
The Complexities Description

Reduct	$IARC$	$GIARC$
	$O(\|C\|^2\|U\|\|U_X\| + \|U_X\|\|C\|\|X_p'\|\|Y_q'\|)$	$O(\|C\|^2\|U\|\|U_X\|)$

According to Theorem 9, one can get that

$$CE_{U\cup U_X}(D|C) = \frac{1}{7^2\times 6}(4^2\times(4-1)\times\frac{1}{12}+3^2 \times(3-1)\times 0)+\frac{4}{49\times 3}=\frac{6}{147}.$$

According to Theorem 10, one can get that $H_{U\cup U_X}(D|C) \approx \frac{1}{7}\times(4\times 0.6+3\times 0)-(-0.086)=0.17$.

6.2 Incremental Algorithms for Adding Multiple Objects

Based on incremental mechanisms of the three entropies, Algorithm 3 introduces a group incremental algorithm for reduct computation based on information entropy.

An example is employed to illustrate Algorithm 3. Similarly, based on complementary entropy, this example updates reduct by using Algorithm 3. And the other two entropies can be used to compute attribute significance in this algorithm in the same way.

Example 5 (Continued from Example 1). Computing new redut based on complementary entropy by using Algorithm 3.

For Table 1, its previous reduct found by using Algorithm 1 based on complementary entropy is $\{c_1, c_2\}$. Suppose that $U_X = \{x_5, x_6, x_7\}$ is the added object set.

According to Step 1, $B = \{c_1, c_2\}$, $U\cup U_X/C = \{\{x_3, x_7\}, \{x_4, x_6\}, \{x_1, x_2\}, \{x_5\}\}$, and $U\cup U_X/B = \{\{x_3, x_7\}, \{x_4, x_6\}, \{x_1, x_2, x_5\}\}$.

Because of $k = 2$ and $k' = 3$, example turns to Step 4.

According to Step 4, we have $E_{U\cup U_X}(D|B) = \frac{6}{49}$ and $E_{U\cup U_X}(D|C) = \frac{4}{49}$. Thus, example needs to add attributes from $C - B$.

In the first loop, $Sig^{outter}_{U\cup U_X}(c_3, B, D) = 0$ and $Sig^{outter}_{U\cup U_X}(c_4, B, D) = \frac{2}{49}$. Thus, $B = B\cup\{c_4\} = \{c_1, c_2, c_4\}$. Now, we have $E_{U\cup U_X}(D|B) = E_{U\cup U_X}(D|C) = \frac{4}{49}$. Example thus stops in Step 4.

According to Step 5, there is no attribute in B need to be deleted and the final reduct is $RED_{U\cup U_X} = \{c_1, c_2, c_4\}$.

The following is the time complexity of above Algorithm 3. As mentioned above, we give in this paper a specific explanation that $|U_X| < |U|$. When a data set is added to the decision tables, according to Theorems 8-10, the time complexity of computing entropy is $O(|U||C| + |U_X||C| + |U_X|^2 + |U||X|)$, and X denotes the object set with identical conditional attribute values in U and U_X. In the algorithm $GIARC$, the time complexity of Step 2 is $O(|C|(|U||U_X||C| + |U||C| + |U_X|^2 + |U||X|)) = O(|C|^2|U||U_X|)$. The time complexity of Step 3 is also $O(|C|^2|U||U_X|)$, and the other steps are constant. So, the total time complexity of algorithm $GIARC$ is $O(|C|^2|U||U_X|)$. When a group of objects are added to a data table, Table 3 shows the time complexities of computing reduct.

In Table 3, we compare the time complexities of $GIARC$ with that of $IARC$, respectively. It is easy to see that, if the size of added object set is very small, i.e., $|U_X|$ is very small, the computational time of $IARC$ is almost identical to that of $GIARC$. However, with the increases of $|U_X|$, especially $|U_X|$ is close to $|U|$, the computational time of $|U_X||C||X_p'||Y_q'|$ is not computationally costless and should not be neglected. Hence, when massive new objects in the databases are generated at once, $GIARC$ is usually more efficient than $IARC$.

7 Experimental Analysis

The objective of the following experiments is to show effectiveness and efficiency of the proposed group incremental algorithm $GIARC$. The data sets used in the experiments are outlined in Table 4, which are all downloaded from UCI repository of machine learning databases. All the experiments have been carried out on a personal

TABLE 4
Description of Data Sets

	Data sets	Samples	Attributes	Classes
1	Breast-cancer-wisconsin(Cancer)	683	9	2
2	Tic-tac-toe	958	9	2
3	Kr-vs-kp	3196	36	2
4	Letter-recognition(Letter)	20000	16	26
5	Krkopt	28056	6	18
6	Shuttle	58000	9	7
7	Person Activity (PA)	164860	8	11
8	Poker-hand	1025010	10	10

computer with Windows 7, Inter(R) Core (TM) i7-2600 CPU (2.66 GHz) and 4.00 GB memory. The software used is Microsoft Visual Studio 2005 and the programming language is C#. And in the data sets, *Shuttle* and *Poker-hand* are preprocessed using the data tool Rosetta.

Eight UCI data sets are employed in the testing. The experiments are divided into three parts, which illustrate effectiveness, efficiency and give a comparison with the existing incremental algorithms, respectively. In the first part, the effectiveness of $GIARC$ is illustrated mainly through comparing it with the classic heuristic attribute reduction algorithm based on information entropy (CAR). In the second part, $IARC$ are first compared with $GIARC$ and the efficiency of $GIARC$ is then illustrated by comparing their computational time. The third part contains the comparison with the existing incremental algorithms. The specific design of experiments for each part is introduced as follows.

7.1 Effectiveness Analysis

In this section, to test the effectiveness of $GIARC$, four common evaluation measures in rough set theory are employed to evaluate the decision performance of the reducts found by CAR and $GIARC$. The four evaluation measures are approximate classified precision, approximate classified quality, certainty measure, and consistency measure, which are shown in Definitions 8 and 9.

In [29] and [30], Pawlak defined the approximate classified precision (AP) and approximate classified quality (AQ) to describe the precision of approximate classification in rough set theory, namely, the discernible ability of a feature subset. If a feature subset has the same AP and AQ with original attributes, this feature subset is considered as has the same discernible ability with original attributes. Hence, this section employs these two measures to estimate the discernible ability of a generated feature subset.

Definition 8. *Let $S = (U, C \cup D)$ be a decision table and $U/D = \{X_1, X_2, \ldots, X_r\}$. The approximate classified precision of C with respect to D is defined as*

$$AP_C(D) = \frac{|POS_C(D)|}{\sum_{i=1}^{r} |\overline{C}X_i|},$$

and the approximate classified quality of C with respect to D is defined as

$$AQ_C(D) = \frac{|POS_C(D)|}{|U|}.$$

In rough set theory, by adopting a reduction algorithm, one can get reducts for a given decision table. Then, based on one reduct, a set of decision rules can be generated from the decision table [29], [35]. Decision rules are used to predict decision values of new objects. Hence, the performance of a set of decision rules may affect its predictive ability. Pawlak [30] introduced two measures to measure the certainty and consistency. However, these two measures cannot give elaborate depictions of the certainty and consistency for a rule set [35]. To evaluate the performance of a rule set, Qian et al. [35] defined certainty measure and consistency measure to evaluate the certainty and consistency of a set of decision rules. And these two measures have attracted considerable attention by many researchers [32], [43], [46]. Hence, α and β are employed to evaluated the decision performance of decision rules induced by the found feature subset in this section.

Definition 9. *Let $S = (U, C \cup D)$ be a decision table, $U/C = \{X_1, X_2, \ldots, X_m\}$, $U/D = \{Y_1, Y_2, \ldots, Y_n\}$, and $RULE = \{Z_{ij} | Z_{ij} : des(X_i) \rightarrow des(Y_j),\ X_i \in U/C, Y_j \in U/D\}$. The certainty measure α of the decision rules on S is defined as*

$$\alpha(S) = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{|X_i \cap Y_j|^2}{|U||X_i|},$$

and the consistency measure β of the decision rules on S is defined as

$$\beta(S) = \sum_{i=1}^{m} \frac{|X_i|}{|U|} \left[1 - \frac{4}{|X_i|} \sum_{j=1}^{n} \frac{|X_i \cap Y_j|^2}{|X_i|} \left(1 - \frac{|X_i \cap Y_j|}{|X_i|} \right) \right].$$

The main objective of this section is to illustrate that $GIARC$ can find a feasible feature subset in a much shorter time, rather than find a more superior one. By comparing with CAR, if discernible ability (evaluated by AP and AQ) and decision performance (evaluated by α and β) of the feature subset found by $GIARC$ are very closed or even identical to that of CAR, then this feature subset can be considered to be feasible. By running algorithms $GIARC$ and CAR on the eight employed data sets, following experiments are to test feasibility and efficiency of $GIARC$.

For each data set in Table 4, 51 percent objects are taken as the basic data set, and the remaining 49 percent objects are taken as incremental objects. When the incremental objects are added to the basic data set, algorithms CAR and $GIARC$ are employed to update reduct of each data set. The experimental results are shows in Tables 5, 6, and 7. These tables show the number of selected features, evaluation results of found feature subsets and computational time of each employed data set. For simplicity, *the number of selected features* is written as NSF.

It is easy to see from Tables 5, 6, and 7 that values of the four evaluation measures of the generated reducts after the updating are very close, and even identical on some data sets. But, the computational time of $GIARC$ is much smaller than that of CAR. In other words, the performance and decision making of the reduct found by $GIARC$ are very close to that of CAR, but $GIARC$ is more efficient. Hence, the experimental results indicate that, compared with the classic reduction algorithm based on entropies CAR, the algorithm $GIARC$ can find a feasible feature subset in a much shorter time.

7.2 Efficiency Analysis

The experimental results in previous section has indicated that $GIARC$ is much more efficient than CAR. In this section, we compare $GIARC$ with $IARC$ to further illustrate the efficiency of algorithm $GIARC$. For each data set in Table 4, let U denote its universe and 51 percent objects ($0.51*|U|$) are selected as the basic data set. Then, we divide the remaining 49 percent objects into five equal parts, denoted by x_i ($|x_i| = \frac{0.49*|U|}{5}, i = 1, 2, \ldots, 5$). Let $X_i = \bigcup_{j=1}^{i} x_i (i = 1, 2, \ldots, 5)$

TABLE 5
Comparison of Evaluation Measures Based on Complementary Entropy

	CAR						$GIARC$					
Data sets	NSF	AQ	AP	α	β	Time/s	NSF	AQ	AP	α	β	Time/s
Cancer	4	1.0000	1.0000	1.0000	1.0000	0.770001	4	1.0000	1.0000	1.0000	1.0000	0.060000
Tic-tac-toe	8	0.9999	1.0000	1.0000	1.0000	2.941168	8	0.9999	1.0000	1.0000	1.0000	0.220000
Kr-vs-kp	29	0.9999	1.0000	1.0000	1.0000	91.35022	29	0.9999	1.0000	1.0000	1.0000	3.381309
Letter	12	0.9999	1.0000	1.0000	1.0000	4564.396	10	0.9999	0.9997	0.9997	0.9994	102.0301
Krkopt	6	0.9999	1.0000	1.0000	1.0000	555.7041	6	0.9999	1.0000	1.0000	1.0000	93.51025
Shuttle	4	1.0000	1.0000	1.0000	1.0000	7913.254	4	1.0000	1.0000	1.0000	1.0000	325.1453
PA	7	1.0000	1.0000	1.0000	1.0000	22220.29	7	1.0000	1.0000	1.0000	1.0000	1022.153
Poker-hand	10	1.0000	1.0000	1.0000	1.0000	868320.6	10	1.0000	1.0000	1.0000	1.0000	62918.36

TABLE 6
Comparison of Evaluation Measures Based on Combination Entropy

	CAR						$GIARC$					
Data sets	NSF	AQ	AP	α	β	Time/s	NSF	AQ	AP	α	β	Time/s
Cancer	4	1.0000	1.0000	1.0000	1.0000	0.820047	4	1.0000	1.0000	1.0000	1.0000	0.066003
Tic-tac-toe	8	0.9999	1.0000	1.0000	1.0000	2.881165	8	0.9999	1.0000	1.0000	1.0000	0.189011
Kr-vs-kp	29	0.9999	1.0000	1.0000	1.0000	36.10806	29	0.9999	1.0000	1.0000	1.0000	1.788102
Letter	12	0.9999	1.0000	1.0000	1.0000	3594.882	11	0.9999	1.0000	1.0000	1.0000	106.2975
Krkopt	6	0.9999	1.0000	1.0000	1.0000	497.5545	6	0.9999	1.0000	1.0000	1.0000	182.5955
Shuttle	4	1.0000	1.0000	1.0000	1.0000	8693.969	4	1.0000	1.0000	1.0000	1.0000	795.0494
PA	7	1.0000	1.0000	1.0000	1.0000	25203.29	7	1.0000	1.0000	1.0000	1.0000	996.1361
Poker-hand	10	1.0000	1.0000	1.0000	1.0000	951264.7	10	1.0000	1.0000	1.0000	1.0000	59039.11

TABLE 7
Comparison of Evaluation Measures Based on Shannon's Entropy

	CAR						$GIARC$					
Data sets	NSF	AQ	AP	α	β	Time/s	NSF	AQ	AP	α	β	Time/s
Cancer	4	1.0000	1.0000	1.0000	1.0000	0.808046	4	1.0000	1.0000	1.0000	1.0000	0.057004
Tic-tac-toe	8	0.9999	1.0000	1.0000	1.0000	2.843163	8	0.9999	1.0000	1.0000	1.0000	0.234013
Kr-vs-kp	29	0.9999	1.0000	1.0000	1.0000	34.59698	29	0.9999	1.0000	1.0000	1.0000	6.522059
Letter	11	0.9998	0.9997	0.9997	0.9997	3671.897	12	0.9999	1.0000	1.0000	1.0000	307.8133
Krkopt	6	0.9999	1.0000	1.0000	1.0000	489.0469	6	0.9999	1.0000	1.0000	1.0000	204.4492
Shuttle	4	1.0000	1.0000	1.0000	1.0000	8512.905	4	1.0000	1.0000	1.0000	1.0000	808.6062
PA	7	1.0000	1.0000	1.0000	1.0000	23183.15	7	1.0000	1.0000	1.0000	1.0000	1063.886
Poker-hand	10	1.0000	1.0000	1.0000	1.0000	865728.3	10	1.0000	1.0000	1.0000	1.0000	60022.59

denotes the incremental group. When each incremental group X_i is added to the basic data set, the two incremental reduction algorithms are used to update the reduct, respectively. The efficiency of the two algorithms are demonstrated by comparing their computational time.

The experimental results are shown in Figs. 1, 2, 3, 4, 5, 6, 7, and 8. In these figures, the y-coordinate pertains to the computational time for updating reduct, and the x-coordinate pertains to the size of incremental group, that is, coordinate value 1, 2, 3, 4, and 5 correspond to adding X_1, X_2, X_3, X_4, and X_5 to the basic data set, respectively. For simplicity, $IARC-L$, $IARC-C$, and $IARC-S$ denote algorithm $IARC$ based on complementary entropy, combination entropy and Shannon's entropy, respectively. Similarly, $GIARC-L$, $GIARC-C$ and $GIARC-S$ denote algorithm $GIARC$ based on the three entropies, respectively.

Figs. 1, 2, 3, 4, 5, 6, 7, and 8 depict the computational time for updating reduct with the two reduction algorithms when different numbers of new objects are added. In view of paper length, for each data set in Table 4, the results of the three entropies are shown in one figure. The experimental results indicate that, in the context of each entropy, $GIARC$ is more efficient than $IARC$ when multiple objects are added to the basic data set. Furthermore, with the number of added objects increasing, for most employed data sets, the efficiency of $GIARC$ is more and more obvious. Hence, the experimental results show that the

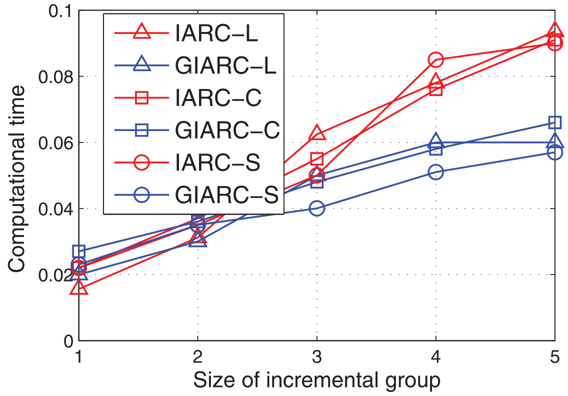


Fig. 1. Cancer.

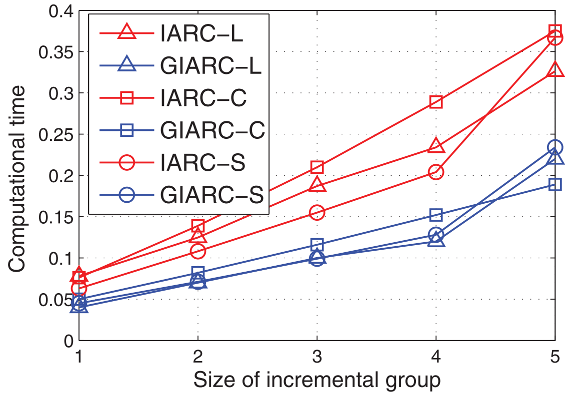


Fig. 2. Tic-tac-toe.

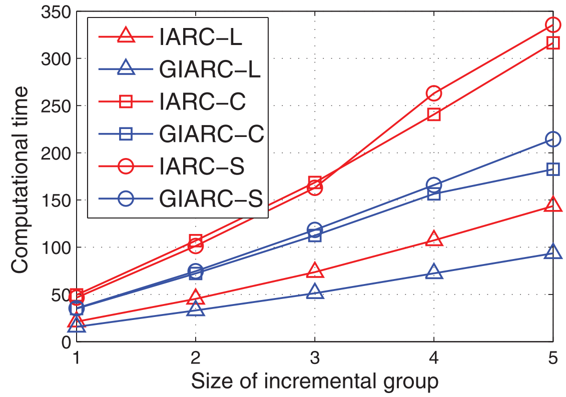


Fig. 3. Kr-vs-kp.

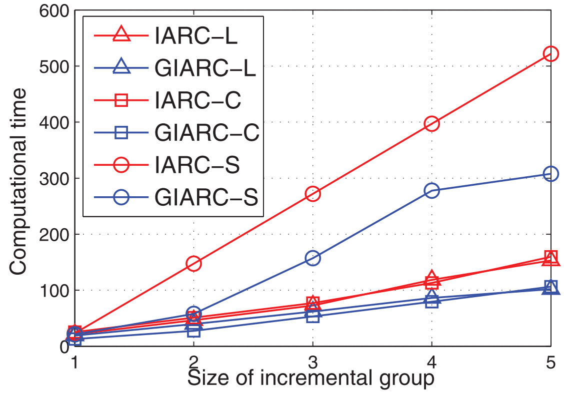


Fig. 4. Letter.

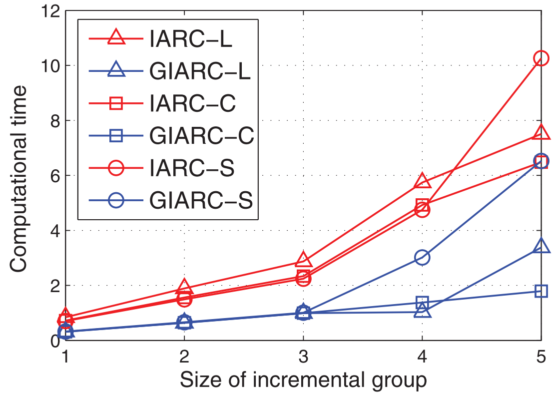


Fig. 5. Krkopt.

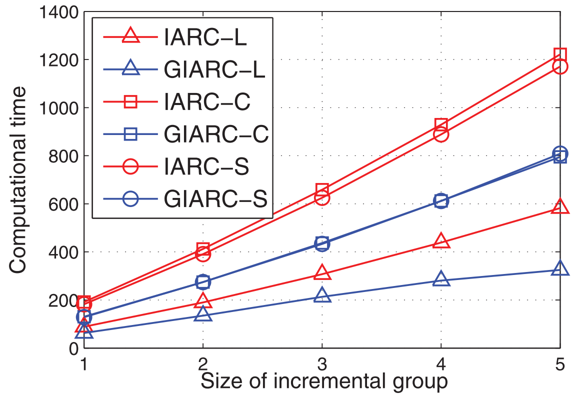


Fig. 6. Shuttle.

group incremental reduction algorithm proposed in this paper is very efficient.

7.3 Comparison with Other Incremental Algorithms

As mentioned in Section 1, there exist in the literature several incremental algorithms for updating redcut. Although an incremental reduction algorithm for finding the minimal reduct was proposed in [25], it is only applicable for information systems without decision attribute. For decision tables, two incremental algorithms were presented in [28] and [41], respectively, whereas both of them are very time consuming. To improve the efficiency, Hu et al. [10] presented an incremental reduction algorithm based on the positive region and showed the experimental results that the algorithm was more efficient than the two algorithms developed in [28] and [41]. Hence, to further illustrate effectiveness and efficiency of algorithm $GIARC$, we compare in this section it with the algorithm in [10]. For convenience, the algorithm in [10] is written as $IRPR$ (incremental reduction based on the positive region) in the following. For each data set in Table 4, 51 percent of the objects are taken as the basic data set, and the remaining 49 percent of the objects are taken as incremental groups. Because Tables 5, 6, and 7 have shown the results of computational time and evaluation measures of $GIARC$, this section only provides in Table 8 the computational time for updating reduct with $IRPR$ and the decision performance of the found reduct.

According to the experimental results in Tables 5, 6, 7, and 8, it is easy to get that the values of the four evaluation measures of the found reducts are very close, and even

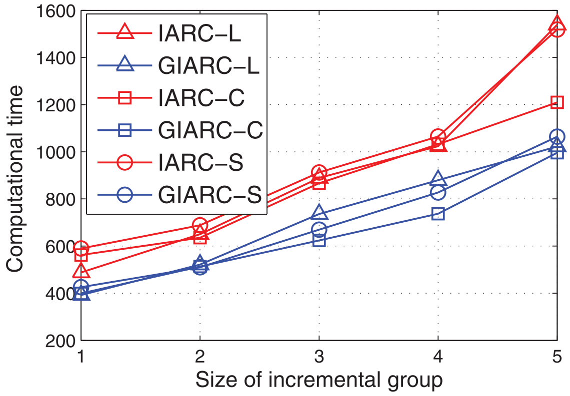


Fig. 7. PA.

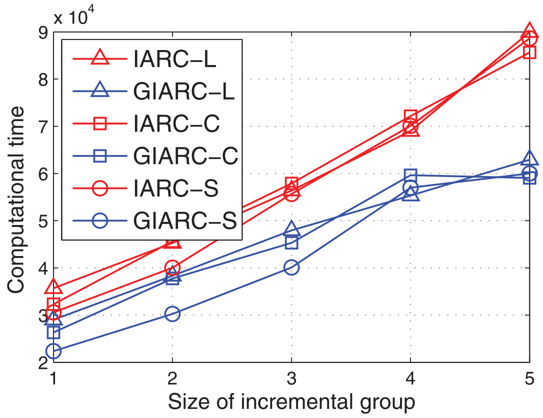


Fig. 8. Poker-hand.

identical on some data sets. But, the computational time of $GIARC$ is much less than that of $IRPR$. In other words, the performance and decision making of the reduct found by $GIARC$ are very close to that of $IRPR$, but $GIARC$ is more efficient. Hence, the experimental results indicate that the algorithm $GIARC$ can find a feasible feature subset in a much shorter time than $IRPR$.

8 Conclusion and Future Work

In this paper, in view of that many real data in databases are generated in groups, an effective and efficient group incremental feature selection algorithm has been proposed in the framework of rough set theory. Compared with existing incremental feature selection algorithms, this algorithm has the following advantages:

1. Compared with classic heuristic feature selection algorithms based on the three entropies, the proposed algorithm can find a feasible feature subset of a dynamically-increasing data set in a much shorter time.
2. When multiple objects are added to a data set, the proposed algorithm is more efficient than existing incremental feature selection algorithms.
3. With the number of added data increasing, the efficiency of the proposed algorithm is more and more obvious.
4. This study provides new views and thoughts on dealing with large-scale dynamic data sets in applications.

Based on above results, some further investigations are as follows:

1. The incremental mechanism of data expanding in groups is in reality the fusion of two data tables. Thus, by generalizing the incremental mechanism, future work would include the information fusion of multidata tables or multigranularity.
2. Further analysis of dynamic data tables shows that the variation of data tables can also include the changes of data values. For data tables with data values changing dynamically, feature selection approaches based on rough set model will be introduced to discover knowledge from dynamic data tables.
3. With the variation of data sets, to predict the decision, the rules extracted from a dynamic data set need to be updated in time. Therefore, it is necessary to devise rules extraction algorithms for a dynamic decision table.

Acknowledgments

This work was supported by National Natural Science Fund of China (No. 71031006 and 61303008), National Key Basic Research and Development Program of China(973) (No. 2011CB311805 and 2013CB329404).

TABLE 8
Computational Time and Evaluation Measures Based on $IRPR$

Data sets	NSA	AQ	AP	α	β	Time/s
Cancer	4	1.0000	1.0000	1.0000	1.0000	0.161009
Tic-tac-toe	8	0.9999	1.0000	1.0000	1.0000	2.388137
Kr-vs-kp	29	0.9999	1.0000	1.0000	1.0000	24.68441
Letter	11	0.9999	1.0000	1.0000	1.0000	334.6461
Krkopt	6	0.9999	1.0000	1.0000	1.0000	200.0974
Shuttle	4	1.0000	1.0000	1.0000	1.0000	903.9317
PA	7	1.0000	1.0000	1.0000	1.0000	2807.636
Poker-hand	10	1.0000	1.0000	1.0000	1.0000	95563.29

REFERENCES

[1] A. An, N. Shan, C. Chan, N. Cercone, and W. Ziarko, "Discovering Rules for Water Demand Prediction: An Enhanced Rough-Set Approach," *Eng. Application and Artificial Intelligence,* vol. 9, no. 6, pp. 645-653, 1996.

[2] W.C. Bang and B. Zeungnam, "New Incremental Learning Algorithm in the Framework of Rough Set Theory," *Int'l J. Fuzzy Systems,* vol. 1, no. 1, pp. 25-36, 1999.

[3] C.C. Chan, "A Rough Set Approach to Attribute Generalization in Data Mining," *Information Sciences,* vol. 107, pp. 177-194, 1998.

[4] H.M. Chen, T.R. Li, D. Ruan, J.H. Lin, and C.X. Hu, "A Rough-Set Based Incremental Approach for Updating Approximations under Dynamic Maintenance Environments," *IEEE Trans. Knowledge and Data Eng.,* vol. 25, no. 2, pp. 274-284, Feb. 2013.

[5] I. Düntsch and G. Gediga, "Uncertainty Measures of Rough Set Prediction," *Artificial Intelligence,* vol. 106, no. 1, pp. 109-137, 1998.

[6] D. Dubois and H. Prade, "Rough Fuzzy Sets and Fuzzy Rough Sets," *Int'l J. General Systems,* vol. 17, pp. 191-209, 1990.

[7] I. Guyon and A. Elisseeff, "An Introduction to Variable Feature Selection," *Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[8] S. Greco, B. Matarazzo, and R. Slowinski, "Rough Sets Theory for Multicriteria Decision Analysis," *European J. Operational Research,* vol. 129, pp. 1-47, 2001.

[9] S. Guo, Z.Y. Wang, Z.C. Wu, and H.P. Yan, "A Novel Dynamic Incremental Rules Extraction Algorithm Based on Rough Set Theory," *Proc. Fourth Int'l Conf. Machine Learning and Cybernetics,* pp. 1902-1907, 2005.

[10] F. Hu, G.Y. Wang, H. Huang, and Y. Wu, "Incremental Attribute Reduction Based on Elementary Sets," *Proc. 10th Int'l Conf. Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, Regina,* pp. 185-193, 2005.

[11] X.H. Hu and N. Cercone, "Learning in Relational Databases: A Rough Set Approach," *Int'l J. Computational Intelligence,* vol. 11, no. 2, pp. 323-338, 1995.

[12] Q.H. Hu, Z.X. Xie, and D.R. Yu, "Hybrid Attribute Reduction Based on a Novel Fuzzy-Rough Model and Information Granulation," *Pattern Recognition,* vol. 40, pp. 3509-3521, 2007.

[13] Q.H. Hu, D.R. Yu, W. Pedrycz, and D.G. Chen, "Kernelized Fuzzy Rough Sets and Their Applications," *IEEE Trans. Knowledge and Data Eng.,* vol. 23, no. 11, pp. 1649-1667, Nov. 2011.

[14] R. Jensen and Q. Shen, "Fuzzy-Rough Sets Assisted Attribute Selection," *IEEE Trans. Fuzzy Systems,* vol. 15, no. 1, pp. 73-89, Feb. 2007.

[15] B. Jerzy and R. Slowinski, "Incremental Induction of Decision Rules from Dominance-Based Rough Approximations," *Electronic Notes in Theoretical Computer Science,* vol. 84, no. 4, pp. 40-51, 2003.

[16] R. Kohavi and G.H. John, "Wrappers for Feature Subset Selection," *Artificial Intelligence,* vol. 97, nos. 1/2, pp. 273-324, 1997.

[17] M. Kryszkiewicz and P. Lasek, "FUN: Fast Discovery of Minimal Sets of Attributes Functionally Determining a Decision Attribute," *Trans. Rough Sets,* vol. 9, pp. 76-95, 2008.

[18] M.Z. Li, B. Yu, O. Rana, and Z.D. Wang, "Grid Service Discovery with Rough Sets," *IEEE Trans. Knowledge and Data Eng.,* vol. 20, no. 6, pp. 851-862, June 2008.

[19] T.R. Li, D. Ruan, W. Geert, J. Song, and Y. Xu, "A Rough Sets Based Characteristic Relation Approach for Dynamic Attribute Generalization in Data Mining," *Knowledge-Based Systems,* vol. 20, no. 5, pp. 485-494, 2007.

[20] J.Y. Liang, C.Y. Dang, K.S. Chin, and C.M. Yam Richard, "A New Method for Measuring Uncertainty and Fuzziness in Rough Set Theory," *Int'l J. General Systems,* vol. 31, no. 4, pp. 331-342, 2002.

[21] J.Y. Liang and Z.Z. Shi, "The Information Entropy, Rough Entropy and Knowledge Granulation in Rough Set Theory," *Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 12, no. 1, pp. 37-46, 2004.

[22] J.Y. Liang, F. Wang, C.Y. Dang, and Y.H. Qian, "An Efficient Rough Feature Selection Algorithm with a Multi-Granulation View," *Int'l J. Approximate Reasoning,* vol. 53, pp. 912-926, 2012.

[23] J.Y. Liang, W. Wei, and Y.H. Qian, "An Incremental Approach to Computation of a Core Based on Conditional Entropy," *Chinese J. System Eng. Theory and Practice,* vol. 4, pp. 81-89, 2008.

[24] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE Trans. Knowledge and Data Eng.,* vol. 17, no. 4, pp. 491-502, Apr. 2005.

[25] Z.T. Liu, "An Incremental Arithmetic for the Samllest Reduction of Attributes," *Acta Electronica Sinica,* vol. 27, no. 11, pp. 96-98, 1999.

[26] J.S. Mi, W.Z. Wu, and X.W. Zhang, "Approaches to Knowledge Reduction Based on Variable Precision Rough Set Model," *Information Sciences,* vol. 159, nos. 3/4, pp. 255-272, 2004.

[27] H.S. Nguyen and D.W. Slezak, "Approximate Reducts and Association Rules Correspondence and Complexity Results," *Proc. Seventh Int'l Workshop New Directions in Rough Sets, Data Mining, and Granular-Soft Computing,* vol. 1711, pp. 137-145, 1999.

[28] M. Orlowska and M. Orlowski, "Maintenance of Knowledge in Dynamic Information Systems," *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory,* R. Slowinski, ed., pp. 315-330, Kluwer Academic Publishers, 1992.

[29] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data.* Kluwer Academic Publishers, 1991.

[30] Z. Pawlak and A. Skowron, "Rudiments of Rough Sets," *Information Sciences,* vol. 177, no. 1, pp. 3-27, 2007.

[31] N. Parthalain, Q. Shen, and R. Jensen, "A Distance Measure Approach to Exploring the Rough Set Boundary Region for Attribute Reduction," *IEEE Trans. Knowledge and Data Eng.,* vol. 22, no. 3, pp. 305-317, Mar. 2010.

[32] J.F. Pang and J.Y. Liang, "Evaluation of the Results of Multi-Attribute Group Decision-Making with Linguistic Information," *Omega,* vol. 40, pp. 294-301, 2012.

[33] Y.H. Qian, J.Y. Liang, W. Pedrycz, and C.Y. Dang, "Positive Approximation: An Accelerator for Attribute Reduction in Rough Set Theory," *Artificial Intelligence,* vol. 174, pp. 597-618, 2010.

[34] Y.H. Qian and J.Y. Liang, "Combination Entropy and Combination Granulation in Rough Set Theory," *Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 16, no. 2, pp. 179-193, 2008.

[35] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, and C.Y. Dang, "Measures for Evaluating the Decision Performance of a Decision Table in Rough Set Theory," *Information Sciences,* vol. 178, pp. 181-202, 2008.

[36] N. Shan and W. Ziarko, "Data-Based Acquisition and Incremental Modification of Classification Rules," *Computational Intelligence,* vol. 11, no. 2, pp. 357-370, 1995.

[37] C.E. Shannon, "The Mathematical Theory of Communication," *The Bell System Technical J.,* vol. 27, nos. 3/4, pp. 373-423, 1948.

[38] D. Slezak, "Approximate Entropy Reducts," *Fundamenta Informaticae,* vol. 53, nos. 3/4, pp. 365-390, 2002.

[39] A. Skowron and C. Rauszer, "The Discernibility Matrices and Functions in Information Systems," *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory,* R. Słowiński, ed. Kluwer Academic Publisher, 1992.

[40] R.W. Swiniarski and A. Skowron, "Rough set Methods in Feature Selection and Recognition," *Pattern Recognition Letters,* vol. 24, pp. 833-849, 2003.

[41] R. Susmaga, "Experiments in Incremental Computation of Reducts," *Rough Sets in Data Mining and Knowledge Discovery,* A. Skowron and L. Polkowski, eds., Springer-Verlag, 1998.

[42] L.Y. Tong and L.P. An, "Incremental Learning of Decision Rules Based on Rough Set Theory," *Proc. World Congress on Intelligent Control and Automation,* pp. 420-425, 2002.

[43] C.Z. Wang, C.X. Wu, D.G. Chen, Q.H. Hu, and C. Wu, "Communicating between Information Systems," *Information Sciences,* vol. 178, pp. 3228-3239, 2008.

[44] G.Y. Wang, H. Yu, and D.C. Yang, "Decision Table Reduction Based on Conditional Information Entropy," *Chinese J. Computer,* vol. 25, no. 7, pp. 759-766, 2002.

[45] W.Z. Wu and W.X. Zhang, "Constructive and Axiomatic Approaches of Fuzzy Approximation Operators," *Information Sciences,* vol. 159, pp. 233-254, 2004.

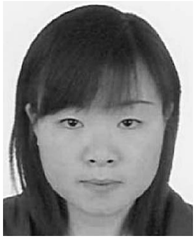
[46] W. Wei, J.Y. Liang, Y.H. Qian, F. Wang, and C.Y. Dang, "Comparative Study of Decision Performance of Decision Tables Induced by Attribute Reductions," *Int'l J. General Systems,* vol, 39, no. 8, pp. 813-838, 2010.

[47] Z.B. Xu, J.Y. Liang, C.Y. Dang, and K.S. Chin, "Inclusion Degree: A Perspective on Measures for Rough Set Data Analysis," *Information Sciences,* vol. 141, pp. 227-236, 2002.

[48] 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 J. Computer,* vol. 29, no. 3, pp. 391-398, 2006.

[49] Y.Y. Yao and Y. Zhao, "Attribute Reduction in Decision-Theoretic Rough Set Models," *Information Sciences,* vol. 178, no. 17, pp. 3356-3373, 2008.
[50] Y.Y. Yao, "Decision-Theoretic Rough Set Models," *Proc. Second Int'l Conf. Rough Sets and Knowledge Technology,* vol. 4481, pp. 1-12, 2007.
[51] M. Yang, "An Incremental Updating Algorithm for Attributes Reduction Based on the Improved Discernibility Matrix," *Chinese J. Computers,* vol. 30, no. 5, pp. 815-822, 2007.
[52] W. Ziarko, "Variable Precision Rough Set Model," *J. Computer and System Science,* vol. 46, no. 1, pp. 39-59, 1993.
[53] A. Zeng, D. Pan, Q.L. Zheng, and H. Peng, "Knowledge Acquisition Based on Rough Set Theory and Principal Component Analysis," *IEEE Intelligent Systems,* vol. 21, no. 2, pp. 78-85, Mar./Apr. 2006.
[54] W. Zhu and F.Y. Wang, "On Three Types of Covering-Based Rough Sets," *IEEE Trans. Knowledge and Data Eng.,* vol. 19, no. 8, pp. 1131-1144, Aug. 2007.
[55] S.Y. Zhao, E.C.C. Tsang, D.G Chen, and X.Z. Wang, "Building a Rule-Based Classifier-a Fuzzy-Rough Set Approach," *IEEE Trans. Knowledge and Data Eng.,* vol. 22, no. 5, pp. 624-638, May 2010.
[56] Z. Zheng and G.Y. Wang, "RRIA: A Rough Set and Rule Tree Based Incremental Knowledge Acquisition Algorithm," *Fundamenta Informaticae,* vol. 59, nos. 2/3, pp. 299-313, 2004.

Jiye Liang received the PhD degree in applied mathematics from Xi'an Jiaotong University. He is a professor of the School of Computer and Information Technology and Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education at Shanxi University. His research interests include feature selection, artificial intelligence, granular computing, data mining, and knowledge discovery. He has published more than 50 articles in international journals.



Feng Wang received the PhD degree from the School of Computer and Information Technology, Shanxi University, China, in 2013. Her research interests include the areas of feature selection, rough set theory, granular computing, and artificial intelligence.

Chuangyin Dang received the PhD degree in operation sresearch/economics from the University of Tilburg, The Netherlands, in 1991. He is an associate professor at the City University of Hong Kong. He is best known for the development of the D1-triangulation of the euclidean space and the simplicial method for integer programming. His current research interests include computational intelligence, optimization theory, applied general equilibrium modeling and computation. He is a senior member of the IEEE and a member of INFORS and MPS.

Yuhua Qian received the MS degree in computers with applications at Shanxi University, China, in 2005. He is an professor of Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, China. He is actively pursuing research in pattern recognition, feature selection, rough set theory, granular computing, and artificial intelligence. He has published more than 20 articles in international journals. He is a member of the IEEE.